# Artificial Intelligence and Democracy

*To my partner Max and my parents*

# Artificial Intelligence and Democracy

## Risks and Promises of AI-Mediated Citizen–Government Relations

Jérôme Duberry

*Senior Researcher, Dusan Sidjanski Centre of Excellence in European Studies, GSI, University of Geneva; Senior Researcher and Lecturer, Albert Hirschman Centre on Democracy, Graduate Institute of International and Development Studies, Switzerland*

Edward Elgar PUBLISHING

Cheltenham, UK • Northampton, MA, USA



Published by
Edward Elgar Publishing Limited
The Lypiatts
15 Lansdown Road
Cheltenham
Glos GL50 2JA
UK

Edward Elgar Publishing, Inc.
William Pratt House
9 Dewey Court
Northampton
Massachusetts 01060
USA

A catalogue record for this book
is available from the British Library

Library of Congress Control Number: 2022934393

This book is available electronically in the Elgaronline
Political Science and Public Policy subject collection
http://dx.doi.org/10.4337/9781788977319

ISBN 978 1 78897 730 2 (cased)
ISBN 978 1 78897 731 9 (eBook)

# Contents

# Acknowledgments

Several people played an important role in the accomplishment of this book. First, I would like to thank Professor Dusan Sidjanski from the University of Geneva, for his intellectual emulation and guidance, as well as the whole team at the Dusan Sidjanski Centre of Excellence in European Studies, for their warm presence and efficient support. At the Global Studies Institute, I would like to thank Professor Nicolas Levrat and the administration team. Inspiring conversations and projects with Dr. Christine Lutringer and the Albert Hirschman Centre on Democracy's team at the Graduate Institute of International and Development Studies, helped me considerably shaping my ideas around technology and citizen participation. Lastly, I would like to thank the Swiss National Science Foundation for supporting this research (Grant 190509). I am also deeply grateful to my partner Max, whose loving presence made this book possible. And of course, to my parents for their unconditional love.

# Introduction

> Strengthening relations with citizens is a sound investment in better policy-making and a core element of good governance. It allows governments to tap new sources of policy-relevant ideas, information and resources when making decisions. Equally important, it contributes to building public trust in government, raising the quality of democracy and strengthening civic capacity. (OECD, 2001, p.1)

What role does artificial intelligence (AI) play in the citizen–government relations? Who is using this technology and for what purpose? How does the use of AI influence power relations in policy-making, and the trust of citizens in democratic institutions? These questions led to the writing of this book. While the early developments of e-democracy and e-participation can be traced back to the end of the 20th century, the growing adoption of smartphones and mobile applications by citizens, and the increased capacity of public administrations to analyze big data, have enabled the emergence of new approaches. Online voting, online opinion polls, online town hall meetings, and online discussion lists of the 1990s and early 2000s have evolved into new generations of policy-making tactics and tools, enabled by the most recent developments in information and communication technologies (ICTs) (Janssen & Helbig, 2018). Online platforms, advanced simulation websites, and serious gaming tools are progressively used on a larger scale to engage citizens, collect their opinions, and involve them in policy processes (Koliba, Zia, & Lee, 2011).

The increasing use of digital technologies in citizen–government relations responds to (1) a demand to digitize public services and make them more efficient, (2) a demand for increased citizen participation in policy-making, and (3) a political will to make public administration data available as discussed below.

First, this adoption of digital technologies by public administrations responds to a growing demand to digitize public action (de Feraudy, 2019). This digital imperative stems from the rapid adoption of digital technologies in Europe. In 2018, it was estimated that 56% of individuals living in the European Union (EU) use social media, and 48% used social media platforms every day or nearly every day in 2019 (Statistica, 2020). According to the Global Web Index, in 2020 internet users aged 16 to 64 were spending an average of two hours and 24 minutes daily on social media.

It is acknowledged, after more than 20 years of research and practice in the field of digital government and the transformation of government through ICT-based reforms, that existing structures will require modification to take complete advantage of the benefits offered by technology (Fountain, 2001; Weerakkody & Dhillon, 2008). For instance, the White Paper on Artificial Intelligence – A European approach to excellence and trust [COM (2020) 65 final] argues that "it is essential that public administrations, hospitals, utility and transport services, financial supervisors, and other areas of public interest rapidly begin to deploy products and services that rely on AI in their activities" (p.65). AI is used for instance by some local governments in the form of an AI-powered social bot to optimize online interaction with citizens and respond to the most common questions. It is thus important to examine how ICTs are transforming policy-making and the relationship between governments and citizens (Janssen & Helbig, 2018).

Second, this digital imperative cumulates with the participatory imperative already weighing on the construction, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017). The idea that encouraging citizen participation can improve the workings of a democracy is echoed in the political analysis of Touraine (1992) who contends that there cannot be any form of democracy without freedom of political choice. This negative concept of democracy and freedom, explained notably by Isaiah Berlin (1969) and Karl Popper (2005), highlights the centrality of citizen participation in democracy.

Citizen participation includes both conventional (e.g. elections) and non-conventional forms of participation (e.g. demonstration). As Rosanvallon and Goldhammer argue

> (…) democratic activity now extends well beyond the framework of electoral-representative institutions (…). The resulting system is complex but, in its own way, coherent. What these various counter-democratic powers have in common is that they describe a new architecture of separated powers and a much more subtle political dynamic than one ordinarily finds in political theory. (Rosanvallon & Goldhammer, 2008, p.249)

Street protests and activism abound in cities around the world and on the internet and social media platforms. Such forms of political participation emerge to express the demands of populations for greater equity, solidarity and to denounce the inaction of politicians on global issues such as climate change. To respond to this new participatory imperative, some governments are offering online and offline participatory instruments (e.g. civic tech, collective intelligence, town hall meetings). This is what Surowiecki (2004) identifies as the wisdom of crowds, while Linders (2012) refers to the shift from e-government to we-government. The latter insists on the new active role that citizens take: they become co-producers of public policies, either in the

form of citizen sourcing (consultation and ideation), government as a platform (information and incentive), or do-it-yourself government (self-organization).

Third, governments make their data available. This is not a new phenomenon: it has been the case for several decades already. What is new, however, is the political will to make all this data available (Harrison et al., 2012). Since 2003, the European Commission (EU) and, in 2009, the American administration, have multiplied their calls for openness (European Commission, 2003; Obama, 2009). These new open data policies aim to make all digital and non-digital government information assets accessible in easy-to-use and digitized formats (Zuiderwijk & Janssen, 2014). Making this publicly funded data available also aims to increase the return on public investment (Arzberger et al., 2004). Open data allows citizens and entities outside of the government to contribute to the policy-making process (European Commission, 2010) and thus become "democratic innovators" (Maier-Rabler & Huber, 2011). This leads to a transfer of data and a transfer of knowledge from inside to outside government (Janssen, Charalabidis, & Zuiderwijk, 2012).

This shift from inside to outside government affects the traditional power relationship between government and the broader environment (Janssen & Helbig, 2018). Because citizens often lack time and expertise, other entities exploit the data made available by public administrations and develop new business models (Zuiderwijk & Janssen, 2014). This represents a risk that these open data policies mainly benefit those who have more resources and expertise, and thus strengthen their argumentation and their own position in the policy-making process.

The increasing use of digital technologies in citizen–government relations presents indeed some challenges. Sometimes perceived as the "go to" solution for the disenchantment of democracy, digital technologies are no panacea of course. Technology was often pursued as an objective in itself, symbolizing modernity more than a desire to really transform participation. E-government and e-participation projects are sometimes based on a certain fetishism of functionalities (e.g. the possibility to "like" contributions) without an a priori needs assessment (Albarède, de Feraudy, Marcou, & Saujot, 2018). What is more, initiatives can be used as a form of veiled rhetoric or as a political marketing strategy for politicians. Many online citizen consultations use closed source or proprietary code platforms with very little or no feedback about the result of the participation (Santini & Carvalho, 2019). Furthermore, behind the participatory processes, other power structures can be hidden (Pickard, 2008) acting in the interest of small groups. Lastly, and maybe most importantly, many citizens lack critical awareness regarding the type of technology used, the actors developing and managing the platform, the actors supporting the initiative, the transparency and accountability of data processing, and questions of cybersecurity and data privacy.

This book is the result of a research project funded by the Swiss National Science Foundation[1] that aimed to explore the opportunities and challenges that AI presents for European liberal democracies. Hence, democracy was our starting point. As we are entering a time of rapid social, economic, political, environmental, and technological transformation, liberal democracy remains the best form of governance, and consequently requires all our attention and care. Inclusive citizen participation based on freedom of opinion and expression are necessary today more than ever to overcome the upcoming global challenges and provide the necessary legitimacy that democratically elected governments need. And when it comes to technologies used in the context of democratic processes, it is important to avoid having the big tech companies make all the decisions. It is up to the populations and their political representatives to decide what role technologies should play in society.

This book considers technology from a (co-)evolutionary innovation studies perspective. By bridging the gap between determinist theories and social constructivism, (co-)evolutionary innovation studies consider technology simultaneously as an active force of change in society, and perceive it as structure, institution, or actor, as well as the result of the design and choices of some social and economic actors (Just & Latzer, 2017). In other words technology can be considered simultaneously as tools and as the outcome of governance (Katzenbach, 2012). This (co-)evolutionary innovation studies framework acknowledges that technology has a form of political agency in society and international relations, and that technology is perpetually developing, with no clear beginning and no end. This approach was adopted by Shah and Kesan (2003, 2010), to highlight the governing role of software, and by Just and Latzer (2017) to examine machine learning algorithms of online platforms.

Moreover, this book considers technology both as a key force of production and a defining mode of social organization and control (Franklin, 2015; Galloway, 2004). With origins in sociology and literary criticism, the Frankfurt School, and more broadly critical theories, aim to shed light on the conditions that enslave people, and seek "to liberate human beings from the circumstances that enslave them" (Horkheimer, 1982, p.244). It contends that social problems are influenced and generated more by societal structures and cultural assumptions than by individual and psychological causes (Geuss, 1981). These theories emerged in relation to social movements, including LGBTIQ+ minorities.[2] In both the broad and the narrow senses, critical theory offers the descriptive and normative bases for social inquiry that seeks to diminish domination and increase freedom in all its dimensions (Bohman, 2021), including in the digital realm (Fuchs, 2021).

The choices made by designers and developers are not neutral and correspond to their view of the world, their culture, their preferences, and their social status. These choices and values are not visible to the user and yet they

influence and shape how technology is used, who can use this technology and for what purpose. Design Justice[3] provides a valuable framework for examining who designs and benefits from technologies, and offers innovative solutions based on participatory design principles for strengthening their appropriation and contributing to their adoption by a wider public, including the populations who have the most disengaged from democratic participation (i.e. poorest and least educated, women, trans folks, B/I/PoC, disabled people, and other marginalized communities). Design Justice aims explicitly to challenge, rather than reproduce, structural inequalities (Costanza-Chock, 2020).

In the context of a geopolitical struggle between autocracy vs. democracy, and where technology and information are weaponized to win the hearts and minds of populations, this book explores how AI mediates the citizen–government relations. Strengthening this relation is crucial as "it contributes to building public trust in government, raising the quality of democracy and strengthening civic capacity" (OECD, 2001, p.1). The relations between governments and citizens span a wide range of interactions at each stage of the policy-making cycle: from problem identification to policy design, through implementation to evaluation. Digital technologies can prove helpful in three main areas: (1) enhancing access to information so that citizens are well informed, (2) enabling citizens to express their views on projects and societal issues that affect them in consultations, (3) engaging citizens in decision-making processes (OECD, 2001). Moreover, many governments responded to the demand of digitizing public action with new e-government services to (a) optimize the effectiveness and efficiency of government services, (b) place the citizen at the center of the design of services rendered by organizations, and (c) increase trust in governments (OECD, 2020).

Among all the digital technologies used, AI has a special place. This book considers AI as "a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions. There are many different technologies that fall under this broad AI definition. At the moment, ML4 techniques are the most widely used" (Craglia et al., 2018, p.18). But AI is also a (1) blurry (i.e. conceptual challenges, ongoing developments and multiple applications), (2) sometimes unreliable (i.e. AI technical or adversarial vulnerabilities, data and algorithm bias), and (3) often opaque (i.e. black box phenomenon) technological agent with (4) various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). While governments are introducing this new technology that offers unprecedented opportunities to increase the efficiency and effectiveness of public action, they must also ensure that it does not contradict core values of liberal democracies.

This book argues that governments may become risk makers when introducing AI in their interactions with citizens, if this introduction is not done according to principles of equality, freedom, and human rights. The risk-taker role differs indeed from the risk-maker in the sense that the decision-maker is the one affected by the consequences of their decision (vs. affecting others). When adopting new technologies, and especially when the new technology is not mature in its development, early adopters may face mistakes, which then may jeopardize the confidence of later adopters in the technology (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). Otherwise, the introduction of AI may change how citizens perceive not only this technology but also their agency and their role in the citizen–government relations.

Tulloch and Lupton (2003) argue that voluntary risk-taking is an "activity in which individuals engage, is perceived by them to be in some sense risky, but is undertaken deliberately and from choice" (pp.10–11). This definition highlights three important elements: (1) reflexivity (or consciousness) that one is taking a risk, (2) capacity (or agency) to make the decision to take the risk, and (3) the voluntary aspect of the decision, which is shaped by social conditions to some extent (Zinn 2015). However, as this book highlights, the AI-mediation of citizen–government relations remains often opaque to the citizen. This leads to question the "voluntary" aspect of the risk-taking role of civil society. And if this use becomes visible, will citizens continue to believe in popular sovereignty once their interactions with the government are systematically mediated by a distrusted version of AI?

Scholars have attempted to explore how and when a society becomes another (Koselleck, 1979; Castoriadis, 1997). Lefort (1988a) examined the transformational role of imaginary in politics and argued that a new political system emerges with the "mutation of the symbolic order" (Lefort, 1986, p.284). For instance, the Enlightenment saw the emergence of individuals autonomous from God, and therefore who could decide for themselves how to organize their collective life without the intermediation or validation of God. This new understanding of the individual led to the transformation of social relations, which "are assumed to be organized, to escape indeterminacy, and to be subject to the will and understanding of human beings" (Lefort, 1988b, p.93). This principle of autonomy (Castoriadis, 1997), and the ontological rupture that it represented in Western Europe, led individuals and society to conceive politics based on principles of equality, freedom, human rights, and the notion of popular sovereignty (Lefort, 1988a).

Popular sovereignty structures the political imaginary of democracy (Diehl, 2019) and forms a "symbolic matrix of democracy" (Lefort, 1986). The principles of equality, freedom, and human rights are the criteria that legitimize political power, and become the normative horizon of democracy (Diehl, 2019). This is illustrated by the French Declaration of Human Rights of 1793,

and more recently, in the European Convention on Human Rights (ECHR). By developing a common understanding of their social existence, social imaginary enables "common practices and a widely shared sense of legitimacy" (Taylor, 2003, p.23) within a nation. But the political imaginary of liberal democracies is in fact constituted of two levels. On the one hand, the normative structure of democracy is settled by a "major imaginary signification that works as the primary reference of democratic representation: popular sovereignty", and on the other hand "social norms manifested through social practices, which are culturally and historically variable" (Diehl, 2019, p.411). These two layers of political imaginary can interact with each other and are interdependent, but do not necessarily have the same temporality. In other words, a political system can simultaneously be grounded in democratic values and tolerate non-democratic practices.

Diehl (2019) highlights these contradictions in democratic societies. On the one hand, liberal democracies are grounded in human rights principles and norms, and on the other hand, they tolerate non-democratic practices for long periods of time after the adoption of such norms, as was the case for women's suffrage. Moreover, this paradox is quite visible at the level of political representations, which at times mix democratic and non-democratic imagery. Diehl (2019) gives the example of the "French mix of revolutionary symbols with the ostentatious style of the monarchy until the present day" (p.410). By distinguishing between these two aspects of the political imaginary, Diehl (2019) contributes to an understanding of how some non-democratic social norms and practices can persist in a democratic system, without necessarily threatening its existence.

As this book illustrates, AI is used in many instances of interaction between citizens and governments. When AI mediates citizens' access to information, citizens' consultations, and their free participation in policy decisions and votes, it introduces a degree of risk and uncertainty that contradicts not only the rationale to use these technologies, but also the political imaginary of democracy. Can European liberal democracies tolerate these practices that erode popular sovereignty without mutating toward a different type of political system? Is AI contributing to "a mutation of the political imaginary" to quote Diehl (2019, p.412)?

The methodological approach followed to achieve the objectives of this book is based on literature review, conceptualization, in-depth case study, and consultation with experts from academia, think tanks, national and sub-national governments, and industry, through workshops, focus groups, and individual interviews. Following the inception phase of the research, which defined the approach to be followed, an exploratory analysis to identify the key challenges and promises for the use of AI in policy-making was conducted. In parallel to this activity, an extensive and multidisciplinary review of the scientific and

gray literature was conducted, which included a literature and policy review, and identifying key research gaps, theoretical frameworks, and real-world use cases. The documents included policy documents from the EU, as well as other international institutions such as the OECD, UNESCO, and World Economic Forum. This review led to the identification of emerging practices in continental Europe (27 EU Member States, United Kingdom, Norway and Switzerland).[4]

After this first step, a series of semi-structured interviews were conducted with 40 experts on the use and impact of AI in policy-making processes. The aim was to gather their views on the promises and challenges of AI in the citizen–government relations. The experts consulted were selected based on their expertise in citizen participation, online platforms, and artificial intelligence. They were identified through literature review and snowballing. A multidisciplinary workshop on the promise of AI for policy-making was held in December 2020 to validate the research findings, and in particular the initial results of the literature review and expert interviews. This workshop first addressed various conceptions of AI (the gap between reality and expectations, education challenges, and media frames). It then considered the promises of AI for fostering citizen mobilization, as well as its pitfalls. It also explored how AI could support collective intelligence processes, including civic tech. It discussed how AI could transform the role and the making of citizens, and finally illustrated key promises of AI for governments (Duberry et al., 2020).

This book is structured as follows. The first chapter presents some key elements for understanding artificial intelligence and its numerous uses. It examines how AI is used to optimize the efficiency and effectiveness of government services. Based on a taxonomy developed by Misuraca and Van Noordt (2020), it shows that AI is increasingly used in the fields of healthcare, education, social and cultural services since it can be considered useful for six types of government challenges: allocating resources, analyzing large datasets, overcoming the shortage of experts, predicting scenarios, managing procedural and repetitive tasks, and diverse data aggregation and summarization. The rest of the book focuses on citizen participation in policy-making.

The second chapter discusses the policy entrepreneurial role of civil society. The Multiple Streams Framework (MSF) is a powerful conceptualization of the policy process, and specifically agenda-setting (Kingdon, [1984], 2011). It argues that policy entrepreneurs need resources (e.g. technology) and specific skills (e.g. engaging multiple audience) to develop and implement tactics (e.g. narrative reframing) through problem, policy and politics streams, to identify and exploit successfully open policy windows. The participation of citizens in policy-making is a direct expression of popular sovereignty. It is based on the assumption that citizens are (1) well informed and are provided with (2)

instance of consultation and (3) decision. The next chapters explore the use of AI in these three areas.

The third chapter examines the use of AI by online platforms, and how it affects access to reliable, relevant and easy-to-find information (OECD). AI enables these platforms to automate information distribution flows, and in particular to rank, filter, and diffuse information. This leads to phenomena such as filter bubble and echo chambers. In this context, the algorithms of social media platforms (in their current development stage) do not benefit civil society and its capacity to make well-information decision. However, social media platforms also offer civil society organizations and social movements an unprecedented opportunity to develop creative advocacy campaigns in order to have their voice heard. They offer a new avenue for civil society to influence policy-making process, or a new policy space according to Leach, Stirling, and Scoones' (2010) definition.

Popular sovereignty is possible when citizen participation is free of any form of coercion, and privacy is secured. Chapter four examines AI-based surveillance tactics and tools from public and private actors. Intelligence services and governments benefit from big data available today. Mass surveillance for national security purposes, as well as digital listening to identify unmet needs in the population are now common practice. Personal data are also collected and then commercialized by a large spectrum of private actors including the Alphabet – Meta digital ads duopoly. By increasing the precision, scale and scope of data collected and processed, these AI-powered surveillance practices also increase citizens' exposure to cybercrimes.

Chapter five explores a world of perpetual political communication and campaigning, where AI enables the automation of digital advertising. Based on the vast amount of data collected by online platforms and other data brokers, one can know with great accuracy how citizens think, what triggers their emotions and decisions. When this "knowledge" is combined with a great ability to reach each individual with a personalized message on a national scale, then one has in one's hands a great power to influence and persuade. The fact that these costly tools are mainly in the hands of governments, political leaders and parties erodes trust and increases an asymmetry of power between citizen and governments. Power lies in the hands of those who hold data and benefit from the AI-powered computational tactics and tools.

Citizens need access to reliable, relevant and easy-to-find information to form their opinion and make political decisions. Chapter six explores AI-based disinformation tactics and tools. Disinformation campaigns target the established trust between citizen and governments, as well as their trust in the information ecosystem itself. They are part of a global power play to reduce the influence of liberal democracies and democratic values in the world. They must also be understood from this global perspective. AI is at the center of this

battlefield: when enabling the diffusion of false news (i.e., by controlling the distribution of content online) and when mitigating their dissemination (i.e., automated content moderation and fact-checking).

The last chapter explores civic technologies. Civic tech refers to technology that aims to increase and deepen democratic participation. They are primarily intended to complement conventional citizen participation and channels of communication previously monopolized by governmental and intergovernmental institutions, as well as address challenges that may be invisible to or neglected by government in a collaborative, problem-centered way. Chapter seven examines AI-powered forms of civic tech. AI is used in this context as well for efficiency purposes: to process a vast number of comments and text published by citizens. It facilitates consulting a larger number of citizens. However, it may be difficult to explain to citizens how AI makes its decisions. In other words, it could make the outcome document suspicious, that is, reducing trust in the process and its perceived legitimacy, as well as hinder citizen participation motivation.

## NOTES

1. Swiss National Science Foundation, Grant 190509, https://data.snf.ch/grants/grant/190509
2. I do not presume the existence of a specific self-conscious and homogenous social community that identifies as "sexual minority." By using the plural "sexual minorities," I am in effect emphasizing that individuals who belong to sexual minorities are rather divided and perceive their bodies and sexuality in many distinct ways.
3. See the website: See designjusticenetwork.org.
4. Close cooperation between these countries in the field of AI: "Declaration of cooperation on AI" adopted by all EU Member States, Norway and Switzerland on 10 April 2018, and the "Coordinated Plan on the Development and Use of Artificial Intelligence Made in Europe" adopted in December 2018, to develop joint actions for stronger and more effective collaboration between the Member States, Norway, Switzerland, and the European Commission in four main areas: boosting investment, making more data available, promoting talent, and building trust.

## REFERENCES

Albarède, M., de Feraudy, T., Marcou, T., and Saujot, M. (2018). *Gouverner et innover dans la ville numérique réelle*. Audacities. IDDRI https://fing.org/wp-content/uploads/2020/02/Audacities_Cas_CivicTechParticipation.pdf [Accessed 25 September 2021]

Aneesh, A. (2006). Virtual Migration. Durham, NC: Duke University Press.

Aneesh, A. (2009). Global labor: Algocratic modes of organization. *Sociological Theory*, 27(4), 347–370.

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L. et al. (2004). An international framework to promote access to data. *Science*, 303(5665), 1777–1778.

Berlin, I. (1969). *Four Essays on Liberty*. Oxford: Oxford University Press.

Bohman, J. (2021). "Critical Theory," *The Stanford Encyclopedia of Philosophy*. Spring 2021 Edition.

Castoriadis, C. (1997). The logic of the magma and the question of autonomy. In: Curtis, D. (ed.) *The Castoriadis Reader.* Oxford: Blackwell Publishing, pp.290–319.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds we Need*. Cambridge, MA: MIT Press.

Craglia M. (Ed.), Annoni A., Benczur P., Bertoldi P., Delipetrev P., De Prato G., Feijoo C., Fernandez Macias E., Gomez E., Iglesias M., Junklewitz H, López Cobo M., Martens B., Nascimento S., Nativi S., Polvora A., Sanchez I., Tolan S., Tuomi I., Vesnic Alujevic L., (2018). *Artificial Intelligence - A European Perspective*, EUR 29425 EN, Publications Office, Luxembourg, ISBN 978-92-79-97217-1, doi:10 .2760/11251, JRC113826.

de Feraudy, T. (2019). Cartographie de la civic tech en France, Observatoire de la civic tech et de la démocratie numérique en France, Décider ensemble.

de Feraudy, T. & Saujot, M. (2017). Une ville plus contributive et durable: Crowdsourcing urbain et participation citoyenne numérique. *Iddri Study*, 4, 1–72.

Diehl, P. (2019). Temporality and the political imaginary in the dynamics of political representation. *Social Epistemology*, 33(5), 410–421.

Duberry, J., Büchi, M., Berryhill, J., Dormeier Freire, A., Garzia, D., Ghernaouti, S., … & Welp, Y. (2020). Promises and pitfalls of artificial intelligence for democratic participation: Workshop proceedings CCDSEE, GSI, University of Geneva, December 10–11, 2020, Virtual Event.

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human–Computer Studies*, 58(6), 697–718.

European Commission. (2003). Directive 2003/98/EC of the European Parliament and of the council of 17 November 2003 on the re-use of public sector information. http://ec.europa.eu/information_society/policy/psi/rules/eu/index_en.htm [Accessed 25 September 2021].

European Commission. (2010). Riding the wave: How Europe can gain from the rising tide of scientific data. Brussels.

European Commission. (2020). *On Artificial Intelligence – A European approach to excellence and trust*. White Paper.

Fountain, J. E. (2001). *Building the Virtual State: Information Technology and Institutional Change*. Washington, DC: Brookings Institution Press.

Franklin, Seb. (2015). *Control. Digitality as Cultural Logic*. Cambridge, MA: MIT Press.

Fuchs, C. (2021). *Social Media: A Critical Introduction*. London: Sage.

Galloway, Alexander R. (2004). *Protocol: How Control Exists after Decentralization*. Cambridge, MA: MIT Press.

Geuss, R. (1981). *The Idea of a Critical Theory: Habermas and the Frankfurt School*. Cambridge: Cambridge University Press.

Harrison, T., Guerrero, S., Burke, G. B., Cook, M., Cresswell, A., Helbig, N. et al. (2012). Open government and e-government: Democratic challenges from a public value perspective. *Information Polity*, 17(2), 83–97.

Horkheimer, M. (1982). *Critical Theory: Selected Essays*. New York: Continuum.

Janssen, M. & Helbig, N. (2018). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*, 35(4), S99–S105.
Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4), 258–268.
Just, N. & Latzer, M. (2017). Governance by algorithms: Reality construction by algorithmic selection on the internet. *Media, Culture & Society*, 39(2), 238–258.
Katzenbach, C. (2012). Technologies as institutions. In: Just, N. and Puppis, M. (eds.), *Trends in Communication Policy Research*. Bristol: Intellect, pp.117–137.
Kingdon, J. (1984). *Agendas, Alternatives and Public Policies*. Boston: Little, Brown.
Kingdon, J. W. (1995). *Agendas, Alternatives, and Public Policies* (2nd edn). New York: Harper Collins College Publisher.
Koliba, C., Zia, A., & Lee, B. (2011). Governance informatics: Utilizing computer simulation models to manage complex governance networks. *The Innovation Journal: Innovations for the Public Sector*, 16(1), 3.
Koselleck, R. (1979). Einleitung. In: Brunner, O., Conze, W,. and Koselleck, R. (eds.), *Geschichtliche Grundbegriffe* (Vol. 1). Stuttgart: Klett-Cotta, pp. viii–xxviii.
Leach, M., Stirling, A. C., & Scoones, I. (2010). *Dynamic Sustainabilities: Technology, Environment, Social Justice*. London: Routledge.
Lefort, C. (1986). The logic of totalitarianism. In: Thompson, J. B. (ed.), *The Political Forms of Modern Society: Bureaucracy, Democracy, Totalitarianism*. Cambridge: Polity Press, pp.273–291.
Lefort, C. (1988a). The question of democracy. In: Lefort, C. (ed.), *Democracy and Political Theory*. Cambridge: Polity Press, pp.9–20.
Lefort, C. (1988b). Interpreting Revolution Within the French Revolution. In Lefort, C. (ed.), *Democracy and Political Theory*. Cambridge: Polity Press, pp.89–114.
Maier-Rabler, U. & Huber, S. (2011). "Open": The changing relation between citizens, public administration, and political authority. *JeDEM*, 3(2), 182–191.
Misuraca, G. & Van Noordt, C. (2020). AI Watch – artificial intelligence in public services: Overview of the use and impact of AI in public services in the EU. *JRC Working Papers* (JRC120399).
Obama, B. (2009). Memorandum for the heads of executive departments and agencies: Transparency and open government. http://www.whitehouse.gov/the_press_of!ce/Transparency_and_Open_Government [Accessed 25 September 2021]
OECD. (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publications.
OECD. (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government. *OECD Public Governance Policy Papers*, No. 02. Paris: OECD Publishing.
Pickard, V. (2008). Cooptation and cooperation: Institutional exemplars of democratic internet technology. *New Media and Society*, 10(4), 625–645.
Popper, K. (2005). *The Logic of Scientific Discovery*. London: Routledge.
Rosanvallon, P. & Goldhammer, A. (2008). *Counter-Democracy: Politics in an Age of Distrust* (Vol. 7). Cambridge: Cambridge University Press.
Santini, R. M. & Carvalho, H. (2019). The rise of participatory despotism: A systematic review of online platforms for political engagement. *Journal of Information, Communication and Ethics in Society*, 17(4), 422–437.
Shah, R. & Kesan, J. (2003). Manipulating the governance characteristics of code. *Info*, 5(4), 3–9.

Shah, R. & Kesan, J. (2010). Software as governance. *Advances of Management Information Systems*, 17, 125.
Statistica. (2020). https://www.statista.com/topics/4106/social-media-usage-in-europe/ .
Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many are Smarter Than the Few and How Collective Wisdom Shapes Business Economies, Societies and Nations*. New York: Doubleday.
Taylor, C. (2003). *Modern Social Imaginaries*. Durham, NC: Duke University Press.
Touraine, A. (1992). *What is Democracy?* UNESCO Courier.
Tulloch, J. & Lupton, D. (2003). Risk and Everyday Life. London: Sage.
We Are Social. (2020). Digital 2020 report. https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media [Accessed 25 September 2021]
Weerakkody, V. & Dhillon, G. (2008). Moving from E-government to T-government: A study of process re-engineering challenges in a UK local authority perspective. *International Journal of Electronic Government Research*, 4(4), 1–16.
Zinn, J. O. (2015). Towards a better understanding of risk-taking: key concepts, dimensions and perspectives. *Health, Risk & Society*, 17(2), 99–114.
Zuiderwijk, A. & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1), 17–29.

# 1. AI to optimize the effectiveness and efficiency of public services

## INTRODUCTION

> With rapid digital technological change, it is inevitable for the government to innovate its traditional methods in order to achieve better citizen engagement, accountability, and interoperability (…). AI can help in freeing up the government labor by its implementation in automating the repetitive tasks resulting in increased transactions speed in the provision of government services and also accurately assessing the outcomes of policy options. AI has a huge potential in different government sectors such as, education, physical infrastructure, transportation, telecommunication, data security and management, finance, healthcare, research and development, policymaking, legal and justice system, etc.
> (Sharma, Yadav, & Chopra, 2020)

Technology and its multiple forms have played a key role in all civilizations. As Hannah Arendt (1958) argued, "[t]ools and instruments are so intensely worldly objects that we can classify whole civilizations using them as criteria" (p.144). Time and specific eras are often identified by the main or the newest technology innovation of that time: we refer to the "digital," "stone," or "steam" ages. Some nations are even characterized by their dominant technological artifacts, such as the Netherlands and windmills, or Japan and micro-electronics (Wyatt, 2008). The role of technology in a society illustrates well the experience of living in a specific era and place (Heilbroner, 1994a, 1994b). This propensity to designate an era or a nation with a technology may be due to the fact that the first scholars to examine in-depth technological change were anthropologists and archaeologists, who use technology as a marker to distinguish eras of development (Mumford, 1961). This role is also indicative of the perception of technology and the choices a nation makes at a given time and in a given place.

Governments have progressively adopted a number of technology innovations to respond to a growing demand to (1) digitalize public action and optimize its operations and services (de Feraudy, 2019), and (2) increase citizen engagement in the development, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017). This is what the concept of e-gov or e-government refers to: using technology to achieve "higher levels of

effectiveness and efficiency in governmental tasks, improvement of processes and procedures, increases the quality of public services, also improves the use of information in the decision-making processes and allows for better communication among different governmental offices" (OAS, n.d.). The e-government efforts were mainly to take advantage of technological advances to (a) optimize the effectiveness and efficiency of government services, (b) put the citizen back at the center of the design of services rendered by organizations, and (c) increase trust in government (OECD, 2020). As Sidjanski (2000) argues, "[t]he emergence of the microcomputer reversed the trend by making it possible to develop horizontal organizations that could to a large extent replace vertical structures" (p.203).

Artificial intelligence (AI) is at the center of a stream of technological solutions, which are increasingly adopted by governments. For instance, it is used to process sensitive information for public health as illustrated by the many applications to combat the spread of the Covid-19 pandemic. AI applications can be considered useful for six types of government challenges: allocating resources, analyzing large datasets, overcoming the shortage of experts, predicting scenarios, managing procedural and repetitive tasks, and diverse data aggregation and summarization (Mehr, Ash, & Fellow, 2017). AI can also provide automated legal advice at lesser cost (Nissan, 2017). However, AI presents numerous challenges, whether they stem from 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). Moreover, AI is often criticized for its black box characteristics: very few experts can understand how the most complex AI systems function, their lines of code evolve with the more data they are fed with (in the case of machine learning algorithms as discussed further), and they are challenging to audit.

This chapter first clarifies the terms artificial intelligence and discusses the conceptual challenges to define this technology. It argues that AI remains this blurry (i.e. conceptual challenges), variable (i.e. ongoing developments and applications), often opaque (i.e. black box phenomenon) agent in the citizen–government relation with various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). It then examines the tasks it can perform and the benefits and risks for governments and other stakeholders to govern with AI. Lastly, it looks at specific uses of AI for public action, and efforts to govern and regulate this technology.

## FROM SYMBOLIC AI TO MACHINE LEARNING

Artificial intelligence is not new. AI has been researched for over 60 years. Its development has taken place over time and through different phases (Darlington, 2017). In his article published in 1950 on computing machines and intelligence, Alan Turing already asked the question of whether machines could think (Tulloch & Lupton, 2003). The Turing test, which is still used today, allows to test an AI when a human being has an interaction with another human being while he thinks he has an interaction with a machine. For many, the Dartmouth Summer Research Project that took place in the summer of 1956 is the birthplace of artificial intelligence (AI). It was during these discussions and exchanges between John McCarthy, Alan Newell, Arthur Samuel, Herbert Simon and Marvin Minsky that AI was conceptualized.

Since then, AI research has developed in stages. During the 1950s until the 1980s, AI research focused on the ambition to make machine think through the use of symbols. This first generation of AI is called symbolic AI, also called "classical AI." John Haugeland (1989) coined the term GOFAI ("Good Old-Fashioned Artificial Intelligence") for symbolic AI. In robotics, the analogous term is GOFR ("Good Old-Fashioned Robotics"). We, humans, make use of symbols to find a specific solution to a mathematical problem. We use the "+" symbol to represent an action, which is adding, and the "=" symbol represents the result of an equation. Similarly to mathematics, we also use symbols to identify the most basic things (e.g. house, or table) and to describe people (e.g. man, woman, doctor, lawyer). We also use symbols to define everyday actions such as walking, drinking, and writing. It is based on the idea that a machine could be trained to think through the use symbols, which represent specific things in the real world (Techslang, 2020). Developers first mapped human "reasoning" to identify rules and symbols we use to think. These symbols are articulated (linked with each other) through a set of rules (e.g. logic, causation). These rules and symbols represent a model of reality, which allow the machine to make a decision by deduction.

Because symbols are necessarily precise representations of reality, they do not allow for implicit knowledge, such as "[a] mother will necessarily be older than her daughter." This is a major limitation in a world that makes extensive use of implicit knowledge. Furthermore, despite obvious reasoning capabilities, the researchers failed to develop the learning capabilities of symbolic AI. Consequently, limited results and less enthusiasm toward this symbolic approach led to the AI Winter: a certain disinterest and reduction in funding for research on this technology in the 1970s.

In the 1990s, computing power and data storage progressed to the point where some complex tasks were feasible for machines. With the emergence of

the internet, the web 2.0, smartphones and social media platforms, new sources of data were soon available. Combined with the increased computing capacity, another approach to AI was possible: statistical approach. These technological advances generated new interest in AI research and attracted new funding (UW, 2006). Statistical AI systems differ from symbolic AI in their inductive process: from a large dataset, they induce trends and create generalizations. For instance, developers can train an AI to recognize cats. To do so, one feeds an AI with a large number of pictures, in which are tagged pictures of cats. The AI will detect patterns and criteria to identify cats in pictures and then create its own definition of cats.

In 1995, Richard Wallace developed the artificial linguistic internet computing entity, which can hold basic conversations. During the same decade, IBM developed the Deep Blue computer that played against Garry Kasparov. In 1996, Deep Blue lost, but won the rematch against Kasparov a year later. Deep Blue had the ability to consider forward six or more steps and could compute 330 million positions per second (Somers, 2013). In 2015, Alphabet DeepMind launched a computer program that can play the game of Go against the best players in the world. AlphaGo is based on an artificial neural network that has been trained on thousands of games played between amateur and professional humans. In 2016, AlphaGo managed to beat Lee Sedol, the best player in the world at the time. Then, the developers let the program play against itself. The result was a new program, AlphaGo Zero, which, through trial and error, managed to beat the original program and all other versions of AlphaGo in 40 days without human intervention or historical data (Silver et al., 2017).

Many AI applications combine both approaches (symbolic and statistical). As an example, natural language processing (NLP) algorithms, which are particularly used by sentiment analysis tactics frequently use a combination of statistical AI (which rely on large amounts of data) and symbolic AI (which consider issues such as grammar rules) (OECD, 2019).

We are currently at the stage of Artificial Narrow Intelligence (ANI). ANI or "applied" AI is developed to solve a specific problem-solving or reasoning task. ANI corresponds to robotized systems and applications that can be considered "intelligent." They cannot mimic human behavior, but they can modestly perform tasks that would require human intelligence, effort, and time to an unsustainable degree, either because of environmental conditions unfavorable to human work or the slowness with which our brains could perform large-scale data analysis (Misuraca & van Noordt, 2020). The most advanced AI systems available today, such as Google's AlphaGo, are still "narrow." Indeed, even if they can to some extent generalize pattern recognition, as for example by transferring the knowledge acquired in the field of image recognition to speech recognition, the human mind remains much more versatile (OECD, 2017).

Two other future steps in the development of AI are worth considering. Artificial superintelligence (ASI) refers to a situation where technology will outperform human intelligence at all times and places, under all conditions and situations. The "technological singularity" refers to that moment in history when human beings are no longer the most intelligent species on earth, but are overtaken by AI. This stage, which for some is more in the realm of science fiction, stirs up dreams and anxieties. Many researchers and ethicists are already trying to prepare our societies for this hypothetical situation, in order to avoid a scenario where AI could take control or even act against the interests of humanity. General AI (AGI) refers to ICT systems with forms of intelligence that are similar to those of humans. Research efforts for this stage are primarily focused on replicating the inner workings of the human brain and applying it to a machine. "AGI would have a strong associative memory and be capable of judgment and decision making. It could solve multifaceted problems, learn through reading or experience, create concepts, perceive the world and itself, invent and be creative, react to the unexpected in complex environments and anticipate" (OECD, 2017). It should be noted, however, that as with the super-intelligence scenario, general AI is far from being realized and may still take decades (or more) to manifest itself (Misuraca & van Noordt, 2020).

This book focuses on the uses of AI that are currently present in our societies (ANI). More futuristic forms of AI (AGI and ASI) will not be discussed due to their future and hypothetical nature. The next section will discuss conceptual challenges to define artificial intelligence.

## CONCEPTUAL CHALLENGES TO DEFINE AI

Despite the excitement around the uses of AI in much of academia, industry, and public institutions, experts fail to agree on one common definition. Artificial intelligence "presents a difficult case for studies of topic sentiment over time" (Fast & Horvitz, 2016, p.963), since this technology is still under development, and its applications are so vast and diverse that there is no general agreement on a common definition of this technology. As Monett and Lewis (2018) argue, "[t]heories of intelligence and the goal of Artificial Intelligence (A.I.) have been the source of much confusion both within the field and among the general public" (p.212). Indeed, AI presents a conceptual challenge that does not enable experts and policy makers to clearly identify its scope of application, its positive and negative consequences.

As Stone et al. (2016) argue, if society approaches AI with fear and suspicion, "missteps that slow AI's development or drive it underground will result, impeding important work on ensuring the safety and reliability of AI technologies" (Stone et al., 2016, p.298). This is particularly problematic in a world where AI is increasingly used in everyday life, including processing

sensitive information for public health. Hence, politicians, regulators, and civil society must acquire a better understanding of this technology (Al-Amoudi & Latsis, 2019) and the associated hopes and concerns it triggers. One must also recommend applying the precautionary principle when the concerns and threats are not fully evaluated and addressed.

First, AI can be considered a field, a discipline, or a science. As McCarthy (1998) states, AI "is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable" (p.2). It can be considered an umbrella term to define a large number of scientific and technological advances. For instance, the Council of Europe (COE) considers AI as "a young discipline of about sixty years, which brings together sciences, theories and techniques (including mathematical logic, statistics, probabilities, computational neurobiology and computer science) and whose goal is to achieve the imitation by a machine of the cognitive abilities of a human being" (Council of Europe, n.d.). First, AI consists of a large area of study in the field of computer science (Fanni, Gabelloni, Alberich-Bayarri, & Neri, 2022). Second, AI is often compared to human capacity. Here, the discussion is about whether AI can mimic a human brain, where it is dedicated to the development of computers capable of engaging in human-like thought processes, including learning, reasoning, and self-correction (Kok, Boers, Kosters, Van der Putten, & Poel, 2009).The relation between AI and human capacity is also well illustrated by the Alan Turing test. This is considered a reliable first test to recognize AI, perhaps because everyone has the ability to conduct one:

> It is quite simple. We place something behind a curtain, and it speaks with us. If we can't make difference between it and a human being, then it will be AI. However, this definition is not formal. Another problem is that this definition does not separate the knowledge from the intellect. (Dobrev, 2004)

AI is indeed often described in relation to human intelligence, or intelligence in general. Indeed, many definitions refer to machines that behave like humans or perform actions that require some form of intelligence (Russel & Norvig, 2010; McCarthy, 2007; Nilsson, 1998; Fogel, 1995; Albus, 1991; Salin & Winston, 1992; McCarthy, 1988; Gardner, 1987, 1983; Newell & Simon, 1976; Bellman, 1978; Minsky, 1969; McCarthy, Minsky, Rochester, & Shannon, 1955/2006). However, these definitions remain vague because of the difficulty of defining and measuring intelligence itself. Thus, this type of definition proposes an ideal target rather than a concrete and measurable research concept.

Third, AI performs a wide range of activities, including "verbal-linguistic, visual-spatial, logical-mathematical, naturalistic, and interpersonal intelligence" (Monett, Lewis, & Thórisson, 2020, p.19). Because AI can "assume some capabilities normally thought to be like human intelligence such as learning, adapting, self-correction" (Mitchell, 2019), it requires an understanding of it through the prism of multiple intelligences: "intelligence is the capacity of an agent to use computation, intended as the capacity to link perception to action in multiple possible sophisticated ways, to increase biological fitness or to accomplish goals" (Monett , Lewis, & Thórisson, 2020, p.19). This leads Wang (2019) to argue that "every working definition of AI corresponds to an abstraction of the human mind that describes the mind from a certain point of view, or at a certain level of abstraction, under the belief that it is what intelligence is really about" (Wang, 2019, p.19).

The many applications in all areas of private and professional life comes from the fact that this technology is a general or foundational technology, just like electricity. The issues associated with this technology are therefore very different from one field to another. It is necessary to distinguish between AI and computer systems that also support human intelligence. Wang (2019) defines AI as "the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources" (p.17). In this working definition, he highlights the combination of two essentials that help distinguish between AI and computer systems: information processing and adaptation. He argues that the information processing capacity of AI consists of choosing and executing tasks, and adjusting its behavior according to its past experiences (Wang, 2019). This operational definition that focuses on technical aspects is close to the one published in February 2020 in the EU White Paper on AI, which describes AI as "a collection of technologies that combine data, algorithms and computing power" (European Commission, 2020). These definitions are consistent with the commonly used definition of AI as "the study of the computations that make it possible to perceive, reason, and act" (Winston, 1992, p.1).

A broader definition is offered by the OECD, which refers to AI as "A machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations or decisions influencing real or virtual environments" (OECD, 2019, p.15). Although very useful, his definition focuses on these two aspects. Thus, it might be too limited to identify the promises and pitfalls of AI in the policy-making process.

In their analysis of the definitions of AI in scientific and gray literature, Samoili et al. (2020) identified four aspects of AI that could be considered as four main features of AI:

- Perception of the environment, which takes into account the complexity of the real world (HLEG, 2019; Nakashima, 1999; Nilsson, 1998; Poole, Mackworth, & Goebel, 1998; Fogel, 1995; Wang, 1995; Albus, 1991; Newell & Simon, 1976).
- Information processing: collection and interpretation of inputs/data (HLEG, 2019; Kaplan & Haenlein, 2019; Nakashima, 1999; Nilsson, 1998; Poole, Mackworth, & Goebel, 1998; Wang, 1995).
- Decision-making, which includes reasoning and learning: taking actions, performing tasks, as well as adapting and reacting to changes in the environment with some level of autonomy (HLEG, 2019; OECD, 2019; Kaplan & Haenlein, 2019; Nilsson, 1998; Poole, Mackworth, & Goebel, 1998; Fogel, 1995; Wang, 1995; Albus, 1991; Newell & Simon, 1976).
- Achieving specific goals: this is regarded as the ultimate reason for the existence of AI systems (HLEG, 2019; OECD, 2019; Kaplan & Haenlein, 2019; Poole, Mackworth, & Goebel, 1998; Fogel, 1995; Albus, 1991; Newell & Simon, 1976).

These key aspects of AI can be found in the operational definition proposed by the High-Level Expert Group on Artificial Intelligence (HLEG):

> Artificial intelligence (AI) refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications) (HLEG, 2018, p.1)

This definition includes indeed the four aspects of AI mentioned previously and identified in the literature. However, it is quite technical. Since this book adopts a social science perspective, the following definitions are most adapted to the objective of this publication:

> AI is a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions. There are many different technologies that fall under this broad AI definition. At the moment, ML4 techniques are the most widely used. (Craglia et al., 2018, p.18)

As discussed in this section, AI is this foundational technology that remains difficult to define precisely. It remains an ongoing project with past, present and future developments. In the definition chosen as a reference for this book, Craglia et al. (2018) highlight key aspects of the agency of AI: its capacity to observe its environment, learn from it, and take smart action or propose decisions. However, this definition also highlights the fact that many technologies fall under the term "AI." In other words, AI remains a conceptual challenge, which makes its understanding and adoption by policy-making stakeholders more challenging as well. Moreover, this conceptual challenge is combined with a level of agency unprecedented in other technologies as discussed in the next section.

## ALGOCRATIC SYSTEM AND AUTONOMOUS TASKS PERFORMED BY AI

As mentioned previously, AI consists of three elements: data, algorithms and computing power. Said differently, an AI system performs three main tasks. First, it collects data from the environment through sensors: it perceives real and/or virtual environments. Second, it builds an abstract model of its environment. Lastly, it produces an output (e.g. recommendations, predictions or decisions).

The environment describes the space that the AI system can observe through perceptions (via sensors) and influence through actions (via actuators). At the core of an AI system lies an abstract representation of the external environment, whether it is a virtual or a real-world environment. This model consists of a set of algorithms (i.e. a set of rules) that represent the structure and/or dynamics of the environment (OECD, 2019). This AI model can be automatically built (also called model building), which means that new data it is fed with improves the precision of its representation of the world (e.g. ML algorithms). It can also be built by human operators and be based on expert knowledge. The model is built according to the type of output it is expected to generate (i.e. objective of the AI system) and performance measures (e.g. accuracy, resources for training, representativeness of the dataset).

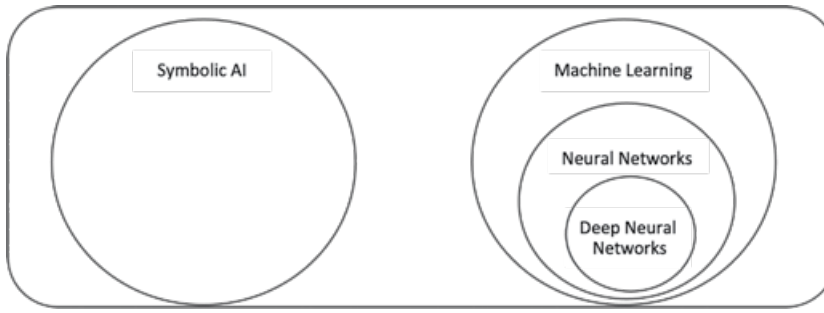
This model allows humans and/or automated tools to derive an outcome such as recommendations, predictions or decisions (also called model inference). In this phase, the AI interprets the raw data collected in relation to the model it has of the environment. Sometimes, the interpretation process can be understood and explained by experts. In some cases, it cannot. This issue is called the black box phenomenon. According to the representativeness of the dataset and the accuracy of the model, the interpretation process will be precise and valid. It can produce one recommendation (e.g. deterministic rules), or several ones (e.g. probabilistic models) (OECD, 2019).

Autonomous driving systems illustrate well how an AI system functions. At its core, the AI model is built from large datasets (e.g. historical driving data, driving rules) and with pre-determined objectives (bring the car safely to a specific destination). Thanks to this abstract representation of the reality, the AI system can (1) perceive its environment (e.g. through sensors such as cameras), (2) make abstract this input data and incorporate it into its model (e.g. object recognition, and location data), and (3) make recommendations in terms of possible short-term futures (model inference). These recommendations have an impact of the environment (e.g. the car accelerates or stops).

Machine Learning (ML) is part of the statistical approach. It is a set of techniques that allow machines to learn in an automated way by detecting and deducing patterns in large datasets rather than by explicit rules and instructions created by a developer. ML approaches frequently teach machines to produce a result by showing them many examples of accurate results. ML can also set a collection of rules and let the machine learn by trial and error. ML can reveal helpful facts to build or adapt an AI model, but it can also be used to interpret the results of an AI model (see Figure 1.1).

ML includes many techniques such as neural networks. The main technology that has enabled the current wave of ML applications is a sophisticated statistical modeling technique called "neural networks." Its deployment is made possible by the constant increase in computational power and the availability of large datasets (also known as "big data"). Neural networks involve the repeated interlinking of thousands or millions of simple transformations into a larger statistical machine, capable of learning complex relationships between inputs and outputs. In short, neural networks change their own code to identify and optimize relationships between inputs and outputs. Deep learning is a sub-category of neural networks. This term refers to especially large neural networks; there is no specific marker to define when a neural network becomes "deep" (OECD, 2019).

In their "AI Watch: Defining Artificial Intelligence" report, Samoili et al. (2020) from the EU Joint Research Centre built a useful AI taxonomy that takes into account political, research and industrial perspectives. It is divided into two main categories: core AI capabilities (e.g. computer vision) and transversal topics (e.g. ethical considerations). As mentioned earlier, AI capabilities can be grouped into two broad categories: (a) reasoning and decision-making, and (b) learning and perception (Table 1.1). The first group includes the transformation of data into knowledge (i.e. transforming data from the real world into information that is understandable and usable by machines), and in so doing, enabling them to make decisions. It includes the AI domains of Reasoning (often through symbolic AI) and Planning. The second group (statistical AI) includes the ability to learn – that is, the ability to extract information and solve problems from the perception of structured or unstructured




*Source:* Adapted from OECD (2019), and the Massachusetts Institute of Technology's (MIT) Internet Policy Research Initiative (IPRI).

*Figure 1.1 Different approach to AI*

data (written and oral language, image, sound, etc.). It also includes the ability to adapt and react to changes and make behavioral predictions among others. It comprises of the domains of learning, communication and perception. In this taxonomy, the categories and sub-categories are related not disjunct. For instance, machine learning is used in computer vision, audio processing and natural language processing (Samoili et al., 2020).

ML and AI are both constituted of sets of algorithms. Algorithms are automated instructions, or step-by-step instructions to process inputs into outputs (Stone, 1972). Today, most algorithms consist of an aggregate of numerous algorithms that function as a computer program (Sandvig, 2014). As Osoba and Welser IV (2017) argue, an algorithm can be defined as "a computable function that can be implemented on computer systems. Machine learning algorithms can also update their behavior in response to experience (input data) and performance metrics" (Osoba and Welser IV, 2017 cited in European Commission, 2020).

Algorithms are now used to govern many aspects of our society and economy (Janssen & Kuk, 2016) as argued by the Committee of Experts MSI-AUT in the 2018 Draft Recommendation of the Committee of Ministers to Member States on the human rights impacts of algorithmic systems entitled "Addressing the Impacts of Algorithms on Human Rights" (Council of Europe, 2018):

> Applications that, often using mathematical optimisation techniques, perform one or more tasks such as gathering, combining, cleaning, sorting, classifying and inferring (ed. personal) data, as well as selection, prioritisation, recommendation and decision-making. Relying on one or more algorithms to fulfil their requirements in the settings in which they are applied, algorithmic systems automate activities in a way that allows the creation of adaptive services at scale and in real time.

*Table 1.1 AI taxonomy*

| | AI Taxonomy | | |
|---|---|---|---|
| Group 1: Transforming data into knowledge. | Reasoning | Ability to represent information about the world in a form that a computer system can employ to resolve complex tasks such as diagnosis of a medical condition or a human oral dialogue (also called Knowledge Representation and Reasoning or KR², KR&R) | |
| | Planning | Ability to solve planning and scheduling problems (i.e. design and execute a sequence of actions where each action has its own set of preconditions to be satisfied before performing the action). Typically performed by intelligent agents, autonomous robots and vehicles. | |
| Group 2: Learning, adapting and reacting to change. | Learning | Ability to automatically learn, decide, predict, adapt to changes and improve without being explicitly programmed. | Machine learning (including neural networks and deep learning). |
| | Communication | Ability to identify, process, understand and/or produce information in human communications (e.g. text). | Natural Language Processing (NLP) (e.g. sentiment analysis, text mining, machine translation). |
| | Perception | Ability to be aware of its environment through sensors. | Computer vision (e.g. face recognition). |
| | | | Audio processing (e.g. speech recognition). |

*Source:* Adapted from AI Watch taxonomy, Samoili et al. (2020), p.17.

This form of delegation of authority to an algorithm that has a decision-making capacity and autonomy is well captured in the notions of "algorithmic regulation" (Yeung, 2017, 2018), "algorithmic governmentality" (Rouvroy, 2015) and algocracy, or algocratic system, originally coined by sociologist Aneesh (2006, 2009), which describes: "[a] governance system in which computer coded algorithms structure, constrain, incentivize, nudge, manipulate or encourage different types of human behavior" (Danaher, forthcoming). Hence, this form of governing with AI (i.e. using AI for public action) presents benefits and risks as discussed in the following section.

## GOVERNING WITH AI: BENEFITS AND RISKS OF AI FOR PUBLIC ACTION

The development of new AI-based initiatives to improve public service delivery is part of an older research tradition. Already in the 1990s, the internet and computer technology helped transform paper-based processes to fully digi-

tized processes and services available online 24/7. The rapid development of internet access and the fast adoption of social media platforms in many liberal democracies generated a growing demand among populations to digitize public action (de Feraudy, 2019). Consequently, governments developed new online services (i.e. e-government) and fostered participation of citizens in some decision-making processes through digital technologies (i.e. e-participation and civic tech).

More recently, governments used artificial intelligence in their relationship with citizens. As Sharma, Yadav, and Chopra (2020) argue, "[w]ith rapid digital technological change, it is inevitable for the government to innovate its traditional methods in order to achieve better citizen engagement, accountability, and interoperability (…)." AI are increasingly used in the fields of healthcare, education, social and cultural services. Moreover, AI can contribute to improving the efficiency and inclusiveness of the policy-making process through optimizing decision-making processes, data and opinion mining, game theory, and agent-based simulation (Milano, O'Sullivan, & Gavanelli, 2014).

> These capabilities and applications could also play a significant role in various governmental tasks related to policy making. For example, and based on the evidence gathered from the case studies reported herein, an early data intelligence exercise can assist public decision makers in detecting emergent societal problems or citizens' concerns much promptly, enabling more timely and accurate policy responses. (Misuraca & van Noordt, 2020, p.19)

Data plays an increasing role in the delivery of public services. Martens (2018) identified three phenomena to consider: the automation and lowering of data collection costs (price effect), the massive increase of available data (quantity effect), and the shift of many face-to-face human activities to the digital domain (substitution effect), have put data sharing at the heart of modern public services and allowed for more efficient and cost-effective delivery. There are both benefits and risks to sharing data. On the one hand, sharing allows for the discovery of new information through the linking of previously unconnected data. On the other hand, thanks to the data collected, it is easier to both know and respond to the needs of the population by adapting the services offered, and to evaluate them. On the other hand, data sharing has its dangers. These include the risk of losing some or all of the data; the possibility of identifying an individual through the combination of many data sources, despite the anonymization of the data; and some negative impacts from the reuse of the data in other contexts to which the owner did not want it disclosed (Involve UK, 2017).

AI research has mainly focused on the governance of AI and to a lesser extent governing with AI. AI differs from previous waves of technology transformation in governments and public organizations. Indeed, AI not only

has the ability to make information available due to its superior computational capacity, but as mentioned previously, it also has the ability to make decisions in place of humans (Latzer & Just, 2020). When AI is further deployed in organizations, this decision-making power can then fundamentally influence how governments and public administrations govern and provide services to citizens (Engstrom, Ho, Sharkey, & Cuéllar, 2020; Mehr, Ash, & Fellow, 2017). Misuraca and van Noordt, in their AI Watch report for the EU Joint Research Centre present a useful taxonomy of AI uses by government. It is based on prior research including Wirtz, Weyerer, and Geyer (2019). This taxonomy is particularly useful because it allowed the two researchers to develop a mapping of AI uses by governments in Europe (EU, UK, Norway and Switzerland).

As shown in Table 1.2, AI presents many benefits for governments to increase the efficiency and effectiveness of their operations and services, such as:

- Improving the knowledge management capacity (e.g. assist in the browsing and finding of relevant data in Slovakia);
- Mapping and predicting risks (e.g. predicts burglaries in Switzerland);
- Automatizing data collection and analysis (e.g. process satellite imagery in Estonia);
- Automatizing data collection and analysis (e.g. process satellite imagery in Estonia); some services (e.g. self-driving snowploughs in Norway), decision-making (e.g. nursery child recruitment system used in Warsaw), and the communication with citizens (e.g. Chatbot to answer frequently asked questions in Latvia).

Although this chapter is not focusing on e-participation and the use of technology to foster the inclusion of civil society in policy making, some of these initiatives also contribute to putting the citizen back at the center of the design of services rendered by the government, such as Natural Language Processing (e.g. AI system to detect the most pressing concerns on Twitter in Ireland). In addition, initiatives that improve the effectiveness of government action may also have the side effect of increasing trust in the capacity of government.

Evidently, citizens benefit strongly from more efficient and effective public action. However, prior research has questioned the real benefits of digitization of government operations and services. As Bannister and Connolly (2020) argue, the promises of digital technologies far exceed the reality and expectations of users (Bannister and Connolly, 2020). Misuraca, Codagnone, and Rossel (2013) and Savoldelli, Codagnone, and Misuraca (2014) have even questioned the merits and real impact of the massive investments in digital that governments have made in recent decades. To what degree do they improve

*Table 1.2 Current and prospective technologies and uses*

| | Type of application | Tasks / Objectives | Example |
|---|---|---|---|
| KNOWLEDGE | AI-empowered Knowledge Management | To create a searchable collection of case descriptions, texts and other insights to be shared with experts for further analysis. | In Slovakia, an AI system is used in the government to assist in the browsing and finding of relevant semantic data. |
| LEARNING | Machine Learning, Deep Learning | While almost all the other categories of AI use some form of Machine Learning, this residual category refers to AI solutions which are not suitable for the other classifications. | In Czechia, AI is used in social services to facilitate 17 citizens to stay in their natural environment for as long as possible. |
| | Predictive Analytics, Simulation and Data Visualization | To identify patterns in data that are consequently used to visualize, simulate or predict new configurations. | Since 2012, the Zurich City Police have been using software that predicts burglaries. Based on these predictions, police could be forwarded to check these areas and limit burglaries from happening. |
| PERCEPTION | Computer Vision and Identity Recognition | Image, video or facial recognition to gain information on the external environment and/or the identity of persons and objects. | In Estonia, the SATIKAS system is in use which is capable of detecting mowed (or the lack of mowed) grasslands on satellite imagery. |
| | Audio Processing | To detect and recognize sound, voices, music and other audio inputs, thus enabling the transcription of spoken words. | Corti in Denmark is used to process the audio of emergency calls in order to detect whether the caller could have a cardiac arrest. |
| | Security Analytics and Threat Intelligence | To analyze and monitoring security information and to prevent or detect malicious activities. | In the Norwegian National Security Authority a new system is used based on machine learning. It is enabling the automatic analysis of any malware detected to improve cybersecurity. |

| | Type of application | Tasks / Objectives | Example |
|---|---|---|---|
| COMMUNICATION | Chatbots, Intelligent Digital Assistants, Virtual Agents and Recommendation Systems | To provide generic advice and behavior-related recommendations to users. | In Latvia, the Chatbot UNA is used to help answer frequently asked questions regarding the process of registering a company. |
| | Natural Language Processing, Text Mining and Speech Analytics | To recognize and analyze speech, written text and communicate back. | In Dublin, an AI system analyzes citizen opinions in the Dublin Region for an overview of their most pressing concerns by analyzing local Twitter tweets with various algorithms. |
| AUTOMATION | Expert and Rule-based Systems, Algorithmic Decision-Making | To facilitate or fully automate decision-making processes of potential relevance. | Nursery child recruitment system used in Warsaw. The algorithm considers data provided by parents during the registration, calculates the score and automatically assigns children into individual nurseries. |
| | Cognitive Robotics, Process Automation and Connected and Automated Vehicles | To automize a process, which can be achieved through robotized hardware or software. | The use of self-driving snowploughs in an airport in Norway in order to improve the clearing of snow on runways. |

*Source:* Adapted from AI Watch (Misuraca & van Noordt, 2020).

the efficiency and effectiveness of government operations and services? Hence, there is a need for auditing the benefits of these burgeoning AI uses in terms of efficiency and effectiveness. Due to the early stages of AI adoption, it is still a difficult task to endeavor.

As mentioned previously, AI presents a conceptual challenge. It remains challenging for many citizens and policy makers to grasp what is AI, what it does, and what its potential benefits and risks are (Duberry & Hamidi, 2021). This is true for AI and for digital technologies at large. The digital divide remains an important issue even in Europe. Indeed, a part of the population is less computerized and less connected than others, in particular the elderly, the countryside, or women. It is both about access to digital technologies and digital skills. In this context, governments' investments to digitize their services can also lead to the exclusion of that part of the population that does not have access or digital skills.

Moreover, AI is characterized by an unprecedented level of agency by structuring, constraining, nudging, and encouraging different types of human behavior (i.e. algocracy, cf. Danaher, forthcoming). To ensure the trust of citizens in this new technology, governments need to ensure that AI-mediated governmental processes and decision-making are transparent and accountable. In other words, citizens need to be able to understand how the decision was

made, and to be able to appeal to the decision. The black box phenomenon describes the difficulty, even for programmers, of dissecting the precise operation of an AI system and more specifically how it arrives at a decision or choice. Since it is constantly adapting its code according to the data provided, its decision process can be particularly difficult to decipher, and just as difficult to audit. This is all the truer since an AI system can be composed of several algorithms. This poses great challenges in terms of legitimacy, transparency, and accountability for decisions made using an AI (Annoni et al., 2018).

A second major concern is the issue of bias in the data used by AI. ML-based applications learn from data. If bias exists in that data, the algorithm will replicate or even reinforce it (Wirtz, Weyerer, & Geyer, 2019). This is particularly the case for historical data, which would lead the algorithm to base itself on a period of history where certain discriminations were widespread, and thus ultimately contribute to reinforcing them. In terms of data, AI can put privacy at risk, especially when the data collected is not voluntarily shared by citizens. This is particularly the case for metadata. This is also the case for sensitive information (e.g. sexual orientation, a health condition) that is inferred from public, non-sensitive data, potentially leading to discriminatory treatment (Floridi, 2017).

Hence, governments face a dilemma. On the one hand, they are asked to improve the performance of their processes and services. To do this, AI presents many opportunities as discussed above. However, they also have the role to protect citizens against the risks that AI presents. It is thus a matter of government responding to two simultaneous demands: governing with AI and governing AI. The following section will discuss some of the main propositions to govern and regulate AI in the context of governmental use.

## EFFORTS AND CHALLENGES TO REGULATE AND GOVERN AI

In recent decades, the progressive digitization of internal processes and public services, as well as the gradual privatization of certain activities previously handled exclusively by the public sector, have required the development of standards and principles of good governance specific to the public sector. The use of AI by the public sector requires a framework of specific standards, principles and values. OECD-SIGMA, in collaboration with the European Commission, has developed a set of principles and a methodological framework for assessing good governance in public administrations.

The values promoted by the European Union (EU) in terms of public service are found in various documents. First, Article 2 of the Treaty on European Union specifies the fundamental values on which the Union is founded. It describes "a society in which pluralism, non-discrimination, tolerance, justice,

solidarity and equality between women and men prevail" (EU, 2008). In addition, there are the rights and freedoms defined by the Charter of Fundamental Rights of the European Union, which only applies when Member States directly implement European regulations or transpose them into their national legislation. In the context of this discussion, we can consider the right to data protection and the right to good administration: "Every person has the right to have his or her affairs handled impartially, fairly and within a reasonable time by the institutions, bodies, offices and agencies of the Union" (Article 41 of the EU Charter of Fundamental Rights).

Specifically for AI, guidelines and other documents developed by universities, think tanks and governmental and non-governmental organizations also provide a number of principles to guide the adoption of AI by governments and public administrations, such as the EU White Paper on AI or the OECD AI principles. The latter proposes five principles that should guide the adoption of AI (OECD, 2020):

- Inclusive growth, sustainable development and well-being
- Human-centered values and fairness
- Transparency and explainability
- Robustness, security and safety
- Accountability.

The OECD AI principles also include five recommendations for policy makers:

- Investing in AI research and development
- Fostering a digital ecosystem for AI
- Shaping an enabling policy environment for AI
- Building human capacity and preparing for labor market transformation
- International cooperation for trustworthy AI.

There are also negative requirements to frame the use of AI in the public sector. The objective of these requirements is indeed to reduce the risk of negative consequences of certain AI applications in public service provision. For example, they require that the AIs used do not have a bias and do not discriminate against a part of the population or a category of citizens. It can also mean requiring a maximum error rate when a government or public administration has delegated a decision-making process to an AI system. These requirements respond to the precautionary principle, in particular for known or anticipated risks. An additional risk with AI comes from its great diversity of applications, and its constant evolution. In other words, it is very difficult today to envisage all the potentially negative consequences, direct and indirect, of using AI in the public sector. This requires principles and standards that are flexible enough to adapt to future situations that are still unknown.

The legal and regulatory framework is of paramount importance to understanding the use of AI in public services. Over the past few decades, a large number of legal and policy tools have been developed to address the growing prominence of AI in the lives of citizens, and in particular the labor market (Frey & Osborne, 2017), health (Jiang et al., 2017), and human rights protection (Eubanks, 2018). AI governance can be defined as "rule-making around algorithms that process data" (Misuraca & van Noordt, 2020, p.49).

The responsibilities associated with the processing of personal data do not fall solely on public sector organizations. When public sector organizations use technologies developed by companies, they make themselves vulnerable to the risk of abuse by those companies, either intentionally or through negligence. Calls for further regulation of online platforms, even after the adoption of the GDPR, demonstrate the growing concern of citizens and civil society organizations about the management and processing of their personal data by technology companies.

Selbst, Boyd, Friedler, Venkatasubramanian, and Vertesi (2019) have shown that efforts to make machine learning algorithms fair (i.e. to ensure that there is no bias or hidden discrimination in the algorithm) tend to "render technical interventions ineffective, inaccurate, and sometimes dangerously misguided when they enter the societal context that surrounds decision-making systems." They identified five different traps or "failing to properly account for or understand the interactions between technical systems and social worlds" (p.59).

- The Framing Trap: "Failure to model the entire system over which a social criterion, such as fairness, will be enforced" (p.60);
- The Portability Trap: "Failure to understand how repurposing algorithmic solutions designed for one social context may be misleading, inaccurate, or otherwise do harm when applied to a different context" (p.61);
- The Formalism Trap: "Failure to account for the full meaning of social concepts such as fairness, which can be procedural, contextual, and contestable, and cannot be resolved through mathematical formalisms" (p.61);
- The Ripple Effect Trap: "Failure to understand how the insertion of technology into an existing social system changes the behaviors and embedded values of the pre-existing system" (p.62);
- The Solutionism Trap: "Failure to recognize the possibility that the best solution to a problem may not involve technology" (p.63).

In terms of AI regulation, the current trend is to treat AI as a technology so specific and unique that it does not fit into existing governance structures, public policies, and laws. A number of organizations and governments have therefore seen the need to produce specific recommendations, strategies and other guidelines for this technology. Their approach, for the most part, shows that they

perceive current norms and governance as inadequate for AI. However, this siloed approach is risky. As Misuraca and van Noordt (2020) argue:

> it would make an enormous difference to think of AI governance as an extension of data protection and competition regulations, acting hand in hand to reduce harms and secure human dignity. Such effort – instead of happening in a vacuum – would help update major existing regulations (i.e. GDPR) to make them work where they do not: by addressing massive imbalances in power, advancing data portability and privacy by design or securing EU wide, public digital infrastructure (p.49).

For example, many existing regulatory frameworks and standards could be applied to AI and its externalities, such as antitrust and consumer protection measures, ethics guidelines, data protection enforcement, intellectual property (IP) protection standards and rules to name just a few. Similarly, both the German Bundeskartellamt and the French Competition Authority in 2019 deemed existing competition laws sufficient to address the challenges posed by the widespread use of AI, and in particular pricing algorithms (Bundeskartellamt & Autorité de la concurrence, 2019).

AI governance is primarily composed of voluntary ethical codes and guidelines. Fjeld, Achten, Hilligoss, Nagy, and Srikumar (2020) mapped these ethical codes developed around the world and identified eight major cross-cutting themes present in most documents they analyzed:

1. Privacy: respect citizens' privacy, both in terms of what type of data is being processed, and in terms of ensuring citizens agency over their personal data;
2. Accountability: existence of accountability mechanisms for the externalities of AI systems;
3. Safety and Security: ensuring that the AI system does not present any vulnerability;
4. Transparency and Explainability: the AI system allows for audit and oversight, including how decisions are made, and where, when, and how they are being used;
5. Fairness and Non-discrimination: ensuring that the design of AI systems and their usage is done according to fairness and inclusivity principles;
6. Human Control of Technology: all important decisions taken by the AI system stay under human review;
7. Professional Responsibility: all professionals and experts involved in the design and maintenance of AI systems follow principles of professionalism and integrity, including the involvement of the stakeholders potentially affected by the AI system;
8. Promotion of Human Values: The purposes to which AI is dedicated, and the means by which it is deployed, must be consistent with our fun-

> damental values (i.e. human rights) and generally promote the welfare of humanity.

When adopting new technologies, and especially when the new technology is not mature in its development, early adopters may face mistakes, which then may jeopardize the confidence of later adopters in the technology (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). It is therefore essential to ensure that AI is not adopted too prematurely for tasks associated with policy making.

## CONCLUDING REMARKS

As discussed in this chapter, governments have progressively adopted a number of technology innovations to respond to a growing demand to (1) digitalize public action and optimize its operations and services (de Feraudy, 2019), and (2) increase citizen engagement in the development, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017).

AI is increasingly deployed by governments to automate and analyze large datasets, enabling the optimization and support of existing processes and services. And yet, this technology remains unregulated and with specific characteristics, which imply a higher degree of uncertainty and risk in the citizen–government relation. AI is indeed this blurry (i.e. conceptual challenges), variable (i.e. ongoing developments and applications), often opaque (i.e. black box phenomenon) agent in the citizen–government relation with various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). In order to ensure that the benefits of this new technology are shared equally, inclusively and transparently among all parts of the population, the need to adopt a human-centric approach to AI as defined by the EU is all the more crucial:

> The Commission has developed key principles to guide the European approach to AI that take into account the social and environmental impact of AI technologies. They include a human-centric way of developing and using AI, the protection of EU values and fundamental rights such as non-discrimination, privacy and data protection, and the sustainable and efficient use of resources. (EU, 2021)

As discussed in the next chapter, conventional forms of citizen participation tend to be in decline, whereas non-conventional forms of participation (i.e. social movements and street protests) have grown significantly in the last decades. Scholars and experts raise a flag about the increase of distrust among citizens toward different forms of public authority, including governments. This is to say that the citizen–government relationship is fragile, and a foundation of any liberal democracy. In this context, one can only recommend

adopting forms of AI where principles of equality, freedom, human rights, and the notion of popular sovereignty are integrated in the technology "by design".

## REFERENCES


Al-Amoudi, I. & Latsis, J. (2019). Anormative black boxes: Artificial intelligence and health policy. In: Al-Amoudi, I. and Lazega, E. (eds.), *Post-Human Institutions and Organizations*. London: Routledge, pp.119–142.

Albus, J. S. (1991). Outline for a theory of intelligence. *IEEE Trans. Systems, Man and Cybernetics*, 21(3), 473–509.

Aneesh, A. (2006). *Virtual Migration*. Durham, NC: Duke University Press.

Aneesh, A. (2009). Global labor: Algocratic modes of organization. *Sociological Theory*, 27(4), 347–370.

Annoni, A., Benczur, P., Bertoldi, P., Delipetrev, B., De Prato, G., Feijoo, C., Fernandez Macias, E., Gomez Gutierrez, E., Iglesias Portela, M., Junklewitz, H., Lopez Cobo, M., Martens, B., Figueiredo Do Nascimento, S., Nativi, S., Polvora, A., Sanchez Martin, J. I., Tolan, S., Tuomi, I., & Vesnic Alujevic, L. (2018). *Artificial Intelligence – A European Perspective*. Luxembourg: Publications Office of the European Union.

Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.

Bannister, F. & Connolly, R. (2020). The future ain't what it used to be: Forecasting the impact of ICT on the public sphere. *Government Information Quarterly*, 37(1), 101410.

Bellman, R. (1978). *An Introduction to Artificial Intelligence: Can Computers Think?* Boston, MA: Boyd & Fraser Publishing.

Bowen, P., Hash, J., & Wilson, M. (2006). *Information Security Handbook: A Guide for Managers*. Gaithersburg: National Institute of Standards and Technology (NIST).

Bundeskartellamt & Autorité de la concurrence. (2019). Algorithms and competition. https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/06_11_2019_Algorithms_and_Competition.html [Accessed 21 August 2021].

Council of Europe (2018). *Addressing the Impacts of Algorithms on Human Rights*. Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence MSI-AUT. https://rm.coe.int/draft-recommendation-of-the-committee-of-ministers-to-states-on-the-hu/168095eecf [Accessed 21 August 2021].

Council of Europe (n.d.). *What is Artificial Intelligence?*. https://www.coe.int/en/web/artificial-intelligence/what-is-ai [Accessed 21 August 2021].

Craglia M. (Ed.), Annoni A., Benczur P., Bertoldi P., Delipetrev P., De Prato G., Feijoo C., Fernandez Macias E., Gomez E., Iglesias M., Junklewitz H, López Cobo M., Martens B., Nascimento S., Nativi S., Polvora A., Sanchez I., Tolan S., Tuomi I., Vesnic Alujevic L., (2018). *Artificial Intelligence - A European Perspective*, EUR 29425 EN, Publications Office, Luxembourg, ISBN 978-92-79-97217-1, doi:10.2760/11251, JRC113826.

Danaher, J. (forthcoming). Freedom in an age of algocracy. In: Vallor, S. (ed.), *Oxford Handbook of Philosophy of Technology*. Oxford: Oxford University Press.

Darlington, K. (2017). *The Emergence of the Age of AI*. OpenMind.

de Feraudy, T. (2019). Cartographie de la civic tech en France, Observatoire de la civic tech et de la démocratie numérique en France, Décider ensemble. https://www

.deciderensemble.com/articles/47158-etude-cartographie-de-la-civic-tech-en-france [Accessed 21 August 2021].
de Feraudy, T. & Saujot, M. (2017). Une ville plus contributive et durable: crowdsourcing urbain et participation citoyenne numérique. *Iddri Study*, 4, 1–72.
Dobrev, D. (2004). *A Definition of Artificial Intelligence*/ Institute of Mathematics and Informatics Bulgarian Academy of Sciences, Sofia, Bulgaria. https://arxiv.org/pdf/1210.1568.pdf [Accessed 21 August 2021].
Duberry, J. & Hamidi, S. (2021). Contrasted media frames of AI during the COVID-19 pandemic: A content analysis of US and European newspapers. *Online Information Review*.
Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human–Computer Studies*, 58(6), 697–718.
Engstrom, D. F., Ho, D. E., Sharkey, C. M., & Cuéllar, M.-F. (2020). Government by algorithm: Artificial intelligence in federal administrative agencies. *SSRN Electronic Journal*.
European Commission. (2020). *White Paper on Artificial Intelligence – A European approach to excellence and trust*. Luxembourg: Publications Office of the European Union. COM(2020) 65 final. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf [Accessed 21 August 2021].
EU (2008). Article 2. *Consolidated Version of the Treaty on European Union*. Official Journal 115, 09/05/2008 P. 0017–0017. Luxembourg: Publications Office of the European Union.
EU (2012). Article 41: *Right to Good Administration*. EU Charter of Fundamental Rights. OJ C 326, 26.10.2012, Luxembourg: Publications Office of the European Union, pp.391–407. https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en [Accessed 21 August 2021].
EU (2021). *AI Excellence: Ensuring that AI Works for People*. Digital Strategy. https://digital-strategy.ec.europa.eu/en/policies/ai-people [Accessed 30 September 2021].
Eubanks, V. (2018). *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press.
Fanni, S. C., Gabelloni, M., Alberich-Bayarri, A., & Neri, E. (2022). Structured reporting and artificial intelligence. In: Fahidi, M. and Pinto dos Santos, D. (eds.), *Structured Reporting in Radiology*. Cham: Springer, pp.169–183.
Fast, E., & Horvitz, E. (2016). *Long-Term Trends in the Public Perception of Artificial Intelligence*. arXiv preprint.
Fjeld, J., Achten, N., Hilligoss, H., Nagy, A., & Srikumar, M. (2020). Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. *Berkman Klein Center Research Publication*, 2020-1.
Floridi, L. (2017). Group privacy: A defence and an interpretation. In: Taylor, L., Floridi, L., and Van der Sloot, B. (eds.), *Group Privacy: New Challenges of Data Technologies* (Vol. 126). Cham: Springer, pp.83–100.
Fogel, D. B. (1995). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. Hoboken, NJ: John Wiley & Sons.
Frey, C. B. & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280.
Gardner, H. (1983). *Frames of Mind; The Theory of Multiple Intelligences*. New York: Basic Books.

Gardner, H. (1987). *The Mind's New Science: A History of the Cognitive Revolution*. New York: Basic Books.
Haugeland, J. (1989). *Artificial Intelligence: The Very Idea*. Cambridge, MA: MIT Press.
Heilbroner, R. (1994a). Technological determinism revisited. In: Smith, M. R. and Marx, L. (eds.), *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge, MA: MIT Press, pp.67–78.
Heilbroner, R. (1994b). Do machines make history?. In: Smith, M. R. and Marx, L. (eds.), *Does Technology Drive History? The Dilemma of Technological Determinism*. Cambridge, MA: MIT Press, pp.53–66.
High Level Expert Group on Artificial Intelligence (HLEG). (2019). *A Definition of AI: Main Capabilities and Disciplines*. Luxembourg: Publications Office of the European Union.
Involve UK. (2017). Better use of data: Balancing privacy and public benefit. https://www.involve.org.uk/sites/default/files/uploads/Better-Use-of-Data-background-briefing.pdf [Accessed 21 August 2021].
Janssen, M. & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371–377.
Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243.
Kaplan, A. & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25.
Kok, J. N., Boers, E. J., Kosters, W. A., Van der Putten, P., & Poel, M. (2009). Artificial intelligence: Definition, trends, techniques, and cases. *Artificial Intelligence*, 1, 270–299.
Latzer, M. & Just, N. (2020). Governance by and of algorithms on the internet: Impact and consequences. In: *Oxford Research Encyclopedia of Communication*, Issue February, Oxford: Oxford University Press, pp. 1–21
Martens, B. (2018). The impact of data access regimes on artificial intelligence and machine learning. *JRC Digital Economy Working Paper*, No. 2018-09.
McCarthy, J. (1988). *The Logic and Philosophy of Artificial Intelligence*. Kyoto Prize Lecture.
McCarthy, J. (1998). *What is Artificial Intelligence?* Computer Science Department, Stanford University.
McCarthy, J. (2007). *What is Artificial Intelligence*. Computer Science Department, Stanford University.
McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955/2006). A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine*, 27(4), 12.
Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. *Ash Center for Democratic Governance and Innovation.* Harvard Kennedy School, August, 1–12. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf [Accessed 21 August 2021].
Milano, M., O'Sullivan, B., & Gavanelli, M. (2014). Sustainable policy making: A strategic challenge for artificial intelligence. *AI Magazine*, 35(3), 22–35.
Minsky, M. L. (1969). *Semantic Information Processing*. Cambridge, MA: MIT Press.

Misuraca, G., Codagnone, C., & Rossel, P. (2013). From practice to theory and back to practice: Reflexivity in measurement and evaluation for evidence-based policy making in the information society. *Government Information Quarterly*, 30, S68–S82.
Misuraca, G., & van Noordt, C. (2020). *Overview of the Use and Impact of AI in Public Services in the EU*. Luxembourg: Publications Office of the European Union.
Mitchell, M. (2019). Artificial intelligence hits the barrier of meaning. *Information*, 10(2), 51.
Monett, D. & Lewis, C.W.P. (2018). *Getting Clarity by Defining Artificial Intelligence – A Survey*. In: Müller, V. (ed.), Conference on Philosophy and Theory of Artificial Intelligence 2017. PT-AI 2017. *Studies in Applied Philosophy, Epistemology and Rational Ethics* (Vol. 44). Springer: Cham. https://doi.org/10.1007/978-3-319-96448-5_21.
Monett, D., Lewis, C. W., & Thórisson, K. R. (2020). Introduction to the JAGI Special Issue "On Defining Artificial Intelligence" – commentaries and author's response. *Journal of Artificial General Intelligence*, 11(2), 1–100.
Mumford, L. (1961). History: Neglected clue to technological change. *Technology & Culture*, 2(3), 230–236.
Nakashima, H. (1999). AI as complex information processing. *Minds and Machines*, 9, 57–80.
Newell, A. & Simon, H. A. (1976). Computer science as empirical enquiry: Symbols and search. *Communications of the ACM*, 19(3), 113–126.
Nilsson, N. J. (1998). *Artificial Intelligence: A New Synthesis*. Burlington, MA: Morgan Kaufmann Publishers.
Nissan, E. (2017). Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement. *AI & Society*, 32(3), 441–464.
OAS (n.d.). About eGov. Portal of the Organization of America States. http://portal.oas.org/portal/sector/sap/departamentoparalagestiónpúblicaefectiva/npa/sobreprogramadeegobierno/tabid/811/default.aspx?language=en-us [Accessed 21 March 2022].
OECD (2017). *OECD Digital Economy Outlook 2017*. Paris: OECD Publishing.
OECD (2019). *Artificial Intelligence in Society*. Paris: OECD Publishing. https://doi.org/10.1787/eedfee77-en [Accessed 21 August 2021].
OECD (2020). The OECD Digital Government Policy Framework: Six dimensions of a Digital Government. *OECD Public Governance Policy Papers*, No. 02. Paris: OECD Publishing.
Osoba, O. A. & Welser IV, W. (2017*). An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica, CA: Rand Corporation.
Poole, D., Mackworth, A., & Goebel, R. (1998). *Computational Intelligence: A Logical Approach*. New York: Oxford University Press.
Rouvroy, A. (2015). Algorithmic governmentality: A passion for the real and the exhaustion of the virtual. *Presentation at the panel "All Watched Over by Algorithms."*
Russel, S. & Norvig, P. (2010). *Artificial Intelligence. A Modern Approach*. London: Pearson.
Salin, E. D. & Winston, P. H. (1992). Machine learning and artificial intelligence: An introduction. *Analytical Chemistry*, 64(1), 49A–60A.
Samoili, S., López Cobo, M., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). *AI Watch. Defining Artificial Intelligence. Towards an*

*operational definition and taxonomy of artificial intelligence*, EUR 30117 EN. Luxembourg: Publications Office of the European Union.
Sandvig, C. (2014). Seeing the sort: The aesthetic and industrial defense of "the algorithm." *Journal of the New Media Caucus*, ISSN, 017X.
Savoldelli, A., Codagnone, C., & Misuraca, G. (2014). Understanding the e-government paradox: Learning from literature and practice on barriers to adoption. *Government Information Quarterly*, 31(SUPPL.1), S63–S71.
Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. *Proceedings of the Conference on Fairness, Accountability, and Transparency* – FAT '19, 59–68.
Sharma, G. D., Yadav, A., & Chopra, R. (2020). Artificial intelligence and effective governance: A review, critique and research agenda. *Sustainable Futures*, 2, 100004.
Sidjanski, D. (2000). *The Federal Future of Europe: From the European Community to the European Union*. Ann Arbor, MI: University of Michigan Press.
Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T., & Hassabis, D. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676), 354–359.
Somers, J. (2013). The man who would teach machines to think. *The Atlantic*, November 2013. https://www.theatlantic.com/magazine/archive/2013/11/the-man-who-would-teach-machines-to-think/309529/ [Accessed 21 August 2021].
Stone, H. S. (1972). *Introduction to Computer Organization and Data Structures*. New York: McGraw-Hill.
Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., … & Teller, A. (2016). Artificial intelligence and life in 2030: The one hundred year study on artificial intelligence. Report of the 2015–2016 Study Panel. Stanford University.
Techslang (2020). *What is Symbolic AI: Examining Its Successes and Failures*. https://www.techslang.com/what-is-symbolic-ai-examining-its-successes-and-failures/ [Accessed 21 August 2021].
Tulloch, J. & Lupton, D. (2003). *Risk and Everyday Life*. London: Sage.
UW (2006). *History of AI*. University of Washington, History of Computing Course (CSEP 590A). https://courses.cs.washington.edu/courses/csep590/06au/projects/history-ai.pdf [Accessed 21 August 2021].
Wang, P. (1995). On the working definition of intelligence. Center for Research on Concepts and Cognition. Indiana University.
Wang, P. (2019). On defining artificial intelligence. *Journal of Artificial General Intelligence*, 10(2), 1–37.
Winston, P. (1992). *Artificial Intelligence*. Reading, MA: Addison-Wesley.
Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector: Applications and challenges. *International Journal of Public Administration*, 42(7), 596–615.
Wyatt, S. (2008). Technological determinism is dead: Long live technological determinism. In: Hackett, E., Amsterdamska, O., Lynch, M., and Wajcman, J. (eds.), *Handbook of Science and Technology Studies*. Cambridge, MA: MIT Press, pp.165–180.
Yeung, K. (2017). "Hypernudge": Big data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118–136.
Yeung, K. (2018). Algorithmic regulation: A critical interrogation. *Regulation & Governance*, 12(4), 505–523.

# 2. Policy entrepreneurs: Skills and resources to identify and exploit open policy windows

## INTRODUCTION

> Democracy and freedom are the exception in the history of humanity, which is marked by authoritarian regimes. Their emergence, their expansion and their future are intertwined with the evolution of Europe and the West. (…) Democracy, freedom, Europe are inseparable.[1]
> (Sidjanski, 1979, p.13)

Touraine (1992) defines democracy as "a regime in which power cannot be taken or held against the will of the majority." This concept of democracy, explained notably by Isaiah Berlin (1969) and Karl Popper (1959), highlights the centrality of citizen participation in democracy, and more particularly its autonomy. At the international level, Article 21 of the Universal Declaration of Human Rights (UDHR) provides that a citizen has the "right to take part in the government of his country, directly or through freely chosen representatives" (Article 21, part 1, UDHR, United Nations, 2001). The same Article 21 (UDHR) also affirms that "[t]he will of the people shall be the basis of the authority of government." More recently and at the regional level, Article 10 of the consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union (EU, 2012) stipulates that the EU Parliament is directly elected by citizens, and that this form of representative democracy constitutes one of the foundations of the Union.

Touraine (1992) contends that there cannot be any form of democracy without freedom of political choice. This is also what Parry, Moyser, and Day (1992) emphasize in their definition of citizen participation: "action by citizens which is aimed at influencing decisions which are, in most cases, ultimately taken by public representatives and officials" (Parry, Moyser, & Day, 1992, p.16). In other words, citizen participation refers to the activities that citizens perform to influence the government (Verba & Nie, 1972). It is based on the idea that some actors designate others to act and speak on their behalf, and where the latter receives "the power of a proxy" (Bourdieu, 1991).

Citizen participation corresponds to "those activities by private citizens that are more or less directly aimed at influencing the selection of governmental personnel and/or the actions they take" (Verba & Nie, 1972, p.2). This definition determines the locus of participation within the sphere of the state, government, or politics at large (Verba, 1967; Verba & Nie, 1972; Verba, Schlozman, & Brady, 1995; Norris, 2002; Bernhagen & Marsh, 2007). It corresponds to the concepts of "formal participation," "conventional modes of participation" (Kaase & Marsh, 1979), "institutional modes of participation" (cf. García-Albacete, 2011; Quintelier & Hooghe, 2013), and "elite-directed action" (Inglehart & Catterberg, 2002). This form of citizen participation has seen a decline in the last several decades in some liberal democracies (Parvin, 2015, 2018), leading some experts to describe this phenomenon as a form of political disengagement.

However, many scholars have questioned this decline by suggesting that citizen participation also comprises other forms of engagement and mobilization. For instance, Jenkins and Carpentier (2013) argue that the participation of civil society in the governing of the state should be conceived in broader terms and include both conventional (e.g. citizen voting in elections) and non-conventional forms of participation (e.g. social movements and street protests). According to this view, conventional citizen participation is a subpart within a broader spectrum of actions performed by civil society to influence the governing of the commons.

Although it remains difficult to precisely define civil society, Wheatley (2010) identifies four key characteristics, namely separation from the state and private capital, self-organization, deliberation, and civility. By being separate from the state and private capital, civil society communicates the demands of the population to the state through advocacy and lobbying initiatives. Thus, civil society takes the role of representing parts of the population and putting pressure on decision-makers to adopt a decision favorable to those they represent. One of the key roles of civil society is to influence policy-making processes.

Conceived in broad terms, the policy-making process is oriented toward the provision of interventions that contribute to the resolution of a large range of perceived societal problems (Stewart Jr, Hedge, & Lester, 2007; Birkland, [2012] 2015; Kolkman, 2020). Civil society can contribute to the policy process through different strategies and tactics, particularly at the agenda-setting stage. The concept of policy entrepreneurship, although not limited to civil society, aims to explain the efforts of some stakeholders to change policy. John Kingdon, a prominent political scientist and keen observer

of Washington DC politics, made the concept of the policy entrepreneur more widely known. He described policy entrepreneur as such:

> Actors who could be in or out of government, in elected or appointed positions, in interest groups or research organizations. But their defining characteristic, much as in the case of a business entrepreneur, is their willingness to invest their resources – time, energy, reputation, and sometimes money – in the hope of a future return. (Kingdon, 1984, p.122)

This chapter discusses key considerations about citizen participation in policy making. These elements will guide the discussion about the use of AI in the subsequent chapters. This chapter starts with the concept of policy entrepreneurship, then focuses on agenda setting and the Multiple Streams Framework (MSF), and finally discusses the emergence of civil society and its role in liberal democracies.

## POLICY ENTREPRENEURSHIP

The relations between government and citizens span a wide range of interactions at each stage of the policy-making cycle: from problem identification to policy design, through implementation to evaluation (OECD, 2001). Conceived in broad terms, the policy-making process is oriented toward the provision of interventions that contribute to the resolution of a large range of perceived societal problems (Stewart Jr, Hedge, & Lester, 2007; Birkland, [2012] 2015; Kolkman, 2020). Policy making has been conventionally viewed as a continuous process that follows a set of steps. This idea is often referred to as the "stage heuristic" (Sabatier, 1999). Lasswell (1951, 1971) identified five stages of policy making: agenda setting, policy formulation, policy adoption, policy implementation and policy evaluation. By setting the agenda, issues are raised to the awareness of policy makers. When on the agenda, policy makers formulate a range of policy options. From these options, an action plan is chosen, enacted into legislation, and adopted. Adoption is followed by implementation of the action plan. The implemented policy is then assessed, and the results of this assessment are brought back to the beginning of the process (Lasswell, 1951, 1971; Easton, 1965; Jones, 1977; Anderson, 1979; Gerston, 2004; Hedge, Lester, & Stewart, 2008; Stewart Jr, Hedge, & Lester, 2007; Birkland, [2012] 2015). Some authors also associate one stage to a specific type of actors (Teisman & van Buuren, 2013) such as civil society organizations for the agenda-setting stage.

The five-step model of policy making conceptualized by Lasswell (1951, 1971) is often criticized for simplifying a process that in reality is not sequential but rather iterative, and for overly reducing the complexity of the

process, particularly by limiting the number of steps to five (Huppe & Hill, 2006). Many variants have been developed, which either increase or decrease the number of steps (Huppe & Hill, 2006). Policy making is an interactive, iterative and complex process that involves and affects many stakeholders and addresses unsolved issues in a large variety of areas (Birkland, [2012] 2015). Other models such as Institutional Analysis and Development framework (Ostrom, 2010), or Experimentalist Governance (Sabel & Zeitlin, 2008, 2012), provide a more accurate depiction of the reality of policy making. Yet this increased realism comes at the expense of additional complexity.

The Multiple Streams Framework (MSF) (Kingdon, 2003) is a valuable resource for gaining an understanding of the policy-making process, and particularly agenda setting, through three separate and independent streams (Knaggård, 2015): problem stream (i.e. identification of the problem), policy stream (i.e. generation and selection of ideas to address this given problem), and politics stream (i.e. political landscape that influences agenda setting) (Danzig, 2018). When designing the Multiple Streams Framework (MSF), Kingdon (1984) had agenda setting in view, which he described as follows: "[t]he separate streams of problems, policies, and politics come together at certain critical times. Solutions become joined to problems, and both of them are joined to favourable political forces" (1984, p.21).

For Kingdon (1995), the policy process is particularly marked by ambiguity. The plurality of different actors' views of the conditions and phenomena under discussion leads to vagueness and confusion (Herweg, Zahariadis, & Zohlnhöfer, 2018; Kingdon, 1995; Zahariadis, 2014). Moreover, fluidity of participation in the policy-making process (i.e. legislative and bureaucratic turnover, and frequent changes in advocacy coalitions); as well as problematic preferences in which policy makers remain uncertain of their objectives or how policies influence them; and unclear technology contribute to the ambiguity of policy making (Fowler, 2019; Herweg, Zahariadis, & Zohlnhöfer, 2018; Kingdon, 1995; Zahariadis, 2014). As Zahariadis (2014) claims about the agenda-setting process:

> [W]e often don't know what the problem is; its definition is vague and shifting. Distinguishing between relevant and irrelevant information is problematic. … Choice becomes less an exercise in solving problems and more an attempt to make sense of a partially comprehensible world … [and] [w]ho pays attention to what and when is critical. (p.28)

Kingdon (1995) intended to conceptualize the ambiguity associated with agenda setting, which is structured around five structural elements: "problems, politics, and policies stream [evolve] independently until policy entrepreneurs

couple streams during open policy windows, leading to agenda setting and decision making" (Fowler, 2019, p.404).

Kingdon (1995) argued that specific circumstances could open what he called policy windows: "opportunit[ies] for advocates of proposals to push their pet solutions, or to push attention to their special problems" (p.165). Policy windows can be triggered by specific events in the problems or politics streams (Fowler, 2019), which stakeholders can then exploit to advance their views and policy proposals (Howlett, 1998). But the emergence of policy window, alone is not enough. It is the combination of the problem, policy and politics streams that will bring together the conditions necessary to put a policy on the agenda (Zahariadis, 2007). It is indeed only when an issue is recognized as a policy problem on the institutional agenda, that the public policy-making process can start addressing it (Béland & Howlett, 2016).

The concept of policy entrepreneurship aims to explain the efforts of some stakeholders to change policy (Mintrom & Norman, 2009, p.658) in various fields such as economy, health, transportation, education, water management, and climate action (Goyal, Howlett, & Chindarkar, 2020). Kingdon (1984) described policy entrepreneurs as such:

> [Actors who] could be in or out of government, in elected or appointed positions, in interest groups or research organizations. But their defining characteristic, much as in the case of a business entrepreneur, is their willingness to invest their resources – time, energy, reputation, and sometimes money – in the hope of a future return. (Kingdon, 1984, p.122)

They play a key role in the policy-making process by designing policy alternatives and coupling them with problems: "[t]he policy entrepreneur works to present a ready package of problems and solutions to policy makers at the right moment. If the policy entrepreneur is successful, the problem will be placed on the political agenda" (Knaggård, 2015, p.450).

Previous research highlighted the capacity of policy entrepreneurs to bring new proposals and ideas on political agenda (Kingdon, 1984, 1995), develop local policy alternatives (Cummings, 2015), generate policy changes (Peters, Jordan, & Tosun, 2017), deliver ambitious policy reform (Aberbach & Christensen, 2014), as well as diffuse policy innovation (Hoyt, 2006; Levi-Faur & Vigoda-Gadot, 2006; Mintrom, 1997; Nay, 2012) in several fields including economic policy (Copeland & James, 2014), education (Verger, 2012), public health (Oliver, 2006), public transportation (Wikström, Eriksson, & Hansson, 2016), and climate action (Krause, 2011; Kwon, Jang, & Feiock, 2014).

Policy entrepreneurship has been examined within a number of theoretical frameworks such as incrementalism, institutionalism, advocacy coalition framework, and punctuated equilibrium (Bakir, 2009; Carter & Jacobs, 2014;

Heikkila et al., 2014; Mintrom & Norman, 2009; Mintrom & Vergari, 1996; Nohrstedt, 2011). However, the Multiple Streams Framework (Kingdon, 1984, 1995, 2003) is a suitable theoretical perspective to examine policy entrepreneurship (Goyal, Howlett, & Chindarkar, 2020) particularly when focusing on the pre-decisional stages of policy making. Hence, many scholars see MSF as primarily, if not only, suited to explaining agenda setting or policy adoption (Herweg, Zahariadis, & Zohlnhöfer, 2018). However, it is also important to note that scholars have added new streams to the MSF, which enables to examine the policy entrepreneurship throughout the policy process (Fowler, 2019; Herweg, Zahariadis, & Zohlnhöfer, 2018; Howlett, McConnell, & Perl, 2015, 2016, 2017; Zahariadis, 2003). The following section will discuss each one of the three streams of MSF in more detail.

## MULTIPLE STREAMS FRAMEWORK

Agenda-setting theory focused primarily on what topics are trending in the news and how these topics and news influence the opinions of audiences (McCombs, Shaw, & Weaver, 2014). It describes the process by which the mass media define what we think and what we are concerned about (Lippmann, 1922). Agenda setting can be divided into two levels: the first one is about making some topics visible to audiences, whereas the second level is about which aspects of this topic are the most important (Littlejohn & Foss, 2010).

Agenda setting also corresponds to the pre-decision phases of policy making (Sidney, 2017). Kingdon (1984, 1995) asserted that agenda setting could be viewed from the perspective of a dynamic between three streams, namely the problem stream, the policy stream, and the politics stream. In this context, policy entrepreneurs, among which civil society, can contribute to "scientific theory building, data production, and publishing, political issue framing, agenda setting, coalition building, business development, marketing and lobbying, management of innovation networks, professional organization" (Voß & Simons, 2014, p.737). This section will discuss each stream separately, and the associated actors and tactics used.

### Problem Stream

Not all problems in society require the involvement of the government. Some problems are best addressed by other actors, whether the private sector, non-profit organizations, or citizens themselves. What's more, not all problems require immediate action. In some cases, citizens tolerate a problem, whether it is because it seems unsolvable or because it does not affect them directly as much, among other reasons. In fact, a problem needs specific characteristics to become a policy problem, namely that citizens find it intolerable, that gov-

ernment can contribute meaningfully, and that it is perceived as legitimate for the government to act and address this problem (Anderson & Shutes, 2014). This is not a static set of conditions since values change in society (e.g. recent changes in perception of climate change). Another situation for problems to become policy problems is when two groups, communities, parts of society compete over resources or power in a specific policy arena (Cobb & Elder, 1983).

Framing is crucial throughout the policy-making process but particularly in the problem stream. Media frame can be defined as "a central organizing idea or story line that provides meaning to an unfolding strip of events (…). The frame suggests what the controversy is about, the essence of the issue" (Gamson & Modigliani, 1987, p.143). This is particularly relevant in times of incertitude, when limited information is available, and when the social media ecosystem is increasingly poisoned with false news. Tuchman (1978) argued that media frames help "organiz[e] everyday reality and the news frame is part and parcel of everyday reality (…). [It] is an essential feature of news" (Tuchman, 1978, p.193). Moreover, media frames structure the news story through origin, forecasts, solutions (Scheufele, 1999) and influence how individuals attribute responsibility (Iyengar, 1991).

The way an issue is framed, either by the government, a political leader, the press, can have a substantial influence on whether it will become a policy problem, and whether the policy-making process will be conducted successfully, from inception, adoption, implementation and to evaluation. Through problem identification and framing, mass media coverage influences which issues the public are aware of, and what their attitudes toward those issues are (Patterson, Semple, Wood, Duffy, & Hilton, 2015). Kitzinger (2004) has shown that public attention is correlated to the way media focus on that issue. In a context of uncertainty, media plays a key role in "shaping public opinion around emerging science and risk issues, and the degree of politicization and polarization of such news coverage may be important and influential factors" (Hart, Chinn, & Soroka, 2020, p.680).

**Actors**

Actors involved in the problem stream tend to be experts in a specific domain relevant to the issue at stake. They are part of a "network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area" (Haas, 1992, p.3), such as for instance science, academia, bureaucracy, and other areas (Goyal, Howlett, & Chindarkar, 2020).

**Tactics**

The main objective of the problem broker (Knaggård, 2015) is to influence the problem definition (Roberts & King, 1991) through various tactics as summarized by Goyal, Howlett, and Chindarkar (2020):

- Raising awareness about the issue at stake (Boasson & Wettestad, 2014; Kalafatis, Grace, & Gibbons, 2015; Knaggård, 2016; Meijerink & Huitema, 2010);
- Linking this problem with other current issues (Brouwer & Huitema, 2018; Mallett & Cherniak, 2018) and well-known events to draw attention and get momentum (Mallett & Cherniak, 2018);
- Changing the perception of the problem by providing new data and indicators (Maor, 2017);
- Reframing existing narratives about the problem (Lovell, 2009; Meijerink & Huitema, 2010);
- Persuade policy makers (Brouwer & Huitema, 2018);
- Delegitimize public institutions and authorities (Goldfinch & Hart, 2003);
- Shop for the most suitable venue to legitimize the problem definition (Baumgartner & Jones, 1993; Brouwer & Huitema, 2018).

**Policy Stream**

The issue is now identified as a policy problem. This stream corresponds to the generation of ideas to address an issue and formulate a new policy. This stream focuses on the following questions: "What is the plan for dealing with the problem? What are the goals and priorities? What options are available to achieve those goals? What are the costs and benefits of each of the options? What externalities, positive or negative, are associated with each alternative?" (Cochran & Malone, 1999, p.46). This stream is about the exploration of the range of possible alternatives for addressing a particular problem. For each alternative, it is a matter of identifying an initial approach to the problem and then designing the specific instruments that constitute this approach. It is then a matter of drafting the legislative or regulatory language for each alternative, in other words of presenting the instruments (i.e. sanctions, subsidies, prohibitions, rights) and explaining the conditions of their application (i.e. to whom, when, how) (Sidney, 2017). Among all these alternatives, a certain number of criteria (feasibility, political acceptability, costs, benefits, real impact, etc.) will make it possible to choose those that will be presented to be adopted.

This stream of the policy-making process is crucial in the sense that it will enable the design of the response to the problem identified. It is both an exploration and a delimitation phase. Exploration because it is a question of exploring all the possibilities. Delimitation because it is also a question of

identifying what is possible. But not only that. This ideation phase also limits the choice proposed to the decision-makers during the following phase. The alternatives considered here must therefore also make it possible to reflect the social, political and economic interests and the power games between the various actors and stakeholders. This may lead to conflicts. As Schattschneider argues it "is the choice of conflicts, and the choice of conflicts allocates power" (1960, p.68).

**Actors**

Actors in the policy stream tend to be also issue experts (Crow, 2010a, 2010b; Oborn, Barrett, & Exworthy, 2011) with field knowledge (Voß & Simons, 2014), and stemming from a range of fields and professional horizons, such as business, consulting, think tanks, public administration, academia, and civil society. Bureaucrats can also take the role of policy entrepreneurs by distributing policy alternatives (Mintrom, 1997; Paquet, 2015).

**Tactics**

Actors in the policy stream can apply several tactics as summarized by Goyal, Howlett, and Chindarkar (2020):

- Provide novel and trustworthy knowledge about the proposal and design options (Anderson, DeLeo, & Taylor, 2020; Braun, 2009; Navot & Cohen, 2015);
- Build good practice models (Mukhtarov & Gerlak, 2013);
- Resort to a "shadow network" to build or experiment with a new idea (Meijerink & Huitema, 2010);
- Launch a trial or pilot project (Brouwer & Huitema, 2018; McFadgen, 2019; Meijerink & Huitema, 2010);
- Take advantage of funding conditions as a donor agency (Meijerink & Huitema, 2010; Mukhtarov & Gerlak, 2013; Shpaizman, Swed, & Pedahzur, 2016);
- Leveraging a window of opportunity for short-term or long-term win (Ugur & Yankaya, 2008);
- Shop for an adequate venue (Mallett & Cherniak, 2018; Meijerink & Huitema, 2010; Shpaizman, Swed, & Pedahzur, 2016);
- Increase the appeal of a particular policy proposal by:
  - Framing it within the prevailing policy paradigm (Béland, 2005);
  - Presenting it as achievable, needed and better than any other alternatives (Brouwer & Huitema, 2018; Goldfinch & Hart, 2003; Palmer, 2015);
  - Tampering with its property or salience (Maor, 2017);
  - Referring to its high valence (Cox & Béland, 2013);

- Associating it with the political agenda (Mukhtarov & Gerlak, 2013);
- Drawing on ethical or professional values (Maor, 2017; Mukhtarov & Gerlak, 2013);
- Suggesting a technically impractical alternative to promote the preferred policy option (Zhu, 2008).

**Politics Stream**

The political landscape has an influence on the agenda-setting process. Politics refers to "the ensemble of practices, discourses and institutions which seek to establish a certain order" (Mouffe, 2000, p.101). The politics stream refers to the political atmosphere regarding which issues are salient and how to balance different interests. It corresponds to the wider environment in which policies are developed. It consists of "factors that influence the body politic, such as swings in national mood, executive or legislative turnover, and interest group advocacy campaigns" (Béland & Howlett, 2016, p.222). The politics stream combines three factors: public opinion, political parties and campaign groups, and administrative and legislative change (Perry & Uuk, 2019). It is about influencing the will or capacity of a government to act on an issue (Goyal, Howlett, & Chindarkar, 2020). In the case of the policy-making process at the level of the European Union, Ackrill, Kay, and Zahariadis (2013) illustrate the politics stream as "[t]he ideological proclivity of incoming governments in EU capitals, the political muscle of bank lobbies in Brussels, and the partisan balance of power in the European Parliament" (Ackrill, Kay, & Zahariadis, 2013, p.873).

Advocacy coalitions and government officials greatly contribute to this stream by framing the public discourse on policy questions (Herweg, Zahariadis, & Zohlnhöfer, 2018; Jenkins-Smith, Nohrstedt, Weible, & Ingold, 2018; Kingdon, 1995; Zahariadis, 2014). When in power, decision-makers tend to follow what the public opinion considers as intolerable and sees positively an action of the government to tackle this issue. The government will organize the agenda in order to stay relevant and popular. Consequently, a change in government, in particular from one side of the political spectrum to the other, will have substantial consequences on the agenda-setting stage (Zahariadis, 2007).

**Actors**

Actors in the politics stream usually consist of members of political parties, politicians, political appointees, and interest groups, amongst other stakeholders (Goyal, Howlett, & Chindarkar, 2020; Mukherjee & Howlett, 2015). They "share a particular belief system – i.e., a set of basic values, causal

assumptions, and problem perceptions – and who show a non-trivial degree of coordinated activity over time" (Sabatier, 1988, p.139).

**Tactics**

Actors in the politics stream can apply several tactics as summarized by Goyal, Howlett, and Chindarkar (2020):

- Become a "prime mover" in the reform journey (Goldfinch & Hart, 2003);
- Raise the chances of a decision in support of the policy alternative by:
  - Politicizing the issue at stake (Hysing, 2009);
  - Engaging public opinion (Roberts & King, 1991);
  - Deploying "salami tactics" (Zahariadis, 2003);
  - Negotiating, bargaining or using side payments (Brouwer & Huitema, 2018; Zahariadis, 2003; Zahariadis & Exadaktylos, 2016);
  - Resorting to keeping a "veil of imprecision" to control the flow of content (Christiansen & Klitgaard, 2010; Christopoulos, 2006);
  - Manipulating the severity or salience of the issue at stake (Boasson & Huitema, 2017; Herweg, Huß, & Zohlnhöfer, 2015).

As discussed in this section, "[p]olicy entrepreneurs are skilled and resourceful actors who couple the three streams together – problems, policies and politics – during open policy windows" (Ackrill, Kay, & Zahariadis, 2013, p.873). The concept of policy space provides another dimension to this discussion. Based on Leach, Scoones, and Stirling's (2010) definition, Prateek, Kumar, Kar, and Krishnan (2021) argue that a policy space is "an avenue (existing and created) to influence policy formulation and implementation by means of formal (legislature, judiciary, quasi-judiciary) and informal (media writings, campaigns, social media advocacy) means" (p.4). Wolmer et al. (2006) relate this concept

> to the extent to which a policy-maker is restricted in decision-making by forces such as the opinions of a dominant actor network or narrative. If there are strong pressures to adopt a particular strategy, a decision-maker may not have much room to consider a wider set of options (Wolmer et al., 2006, p.13).

Prateek, Kumar, Kar, and Krishnan (2021, p.4) identified five policy spaces and their corresponding strategies (adapted from Leach, Scoones, & Stirling, 2010, p.138):

- Bureaucratic spaces: lobbying with state representatives, public officers, and bureaucrats who take part in policy-making processes;
- Invited spaces: participating in meetings and workshops led by governments to share views, influence the agenda, and propose alternative options.

- Conceptual spaces: introducing new ideas through academic papers, online and print publications, reports, and bulletins;
- Practical space: providing relevant facts, scientific data and evidence to policy makers from field assessment, pilot projects, and case study;
- Popular space: Leading or contributing to grassroot movements, online and offline protests, petitions, boycotts (i.e. non-conventional forms of political participation).

The bureaucratic and invited spaces exist formally and explicitly to accommodate the views of stakeholders in the policy-making process, while the other three (conceptual, practical and popular spaces) are created by stakeholders themselves to make their voices heard. However, their success depends on skills, resources, and strategies (Fowler, 2019). The following section will discuss the main skills that enable them to contribute successfully to the policy-making process.

## SEVEN KEY SKILLS OF POLICY ENTREPRENEURS

Previous research explored the criteria to determine what constitutes policy entrepreneurs. Mintrom and Norman (2009) determined specific skills for policy entrepreneurs, such as defining problems, building teams, and leading by example. Arklay, Van Acker, and Hollander (2018) explored the strategies developed by policy entrepreneurs, while Aviram, Cohen, and Beeri (2020) identified certain psychological traits such as social acuity, persuasion, and trust building among "government leaders, bureaucrats, politicians, NGO advocates, private-sector stake-holders and advocates, policy consultants, interest group activists, or social issue campaigners" (p.37).

Mintrom (2019) identified seven key skills for policy entrepreneurs: strategic thinking, team building, collecting evidence, making arguments, engaging multiple audiences, negotiating, and networking. These skills in conjunction with particular attributes enable policy entrepreneurs to deploy strategies and tactics (Mintrom, 2019).

1. Strategic thinking can be described "as a way of solving strategic problems that combines a rational and convergent approach with creative and divergent thought processes" (Bonn, 2005, p.337). It consists of deciding on a specific goal and then establishing the set of actions they will have to take and the necessary resources they will require to reach that objective (Mintrom, 2019). Strategic thinking is also action oriented. As Mintzberg, Ahlstrand, and Lampel (1998, p.42) claimed "there are times when thought should precede action, and guide it (…). Other times, however, especially during or immediately after major unexpected shifts in the environment,

thought must be so bound up with action that 'learning' becomes a better notion than 'designing' for what has to happen. And then, perhaps most common are a whole range of possibilities in between, where thought and action respond to each other." This is what Weick (1983, p.225) described as the capacity "act thinkingly." It is particularly useful to overcome ambiguity and make sense of a complex world, as well as respond to situations and conditions that are continuously changing (Dixit & Nalebuff, 2008). In the context of policy entrepreneur, Kalil (2017) stressed the need to start with a well-defined objective, to have a sizeable and expanding "toolbox" of policy options that can be applied to solve specific issues, and to build strong relationships with others and minimize the barriers to supporting your policy options.

2. Policy entrepreneurs need to be team players. While individual people are frequently the drivers of change, their force does not stem solely from the strength of their own ideas. As Petridou (2014) claimed, "entrepreneurial actions are carried out by teams and not just one heroic, lonely individual" (p.S22). To achieve their objective, they must count on the support and the active engagement of local policy contexts (Kingdon, 1984; Mintrom & Salisbury, 2014; Rabe, 2004). It allows them to better capture what motivates and concerns who they need those they need to convince (Mintrom, 2019).
3. Collecting evidence is a crucial skill for policy entrepreneurs. Evidence can be used to shed light on the issue and support one specific policy option (Kingdon, 1984). In this context, policy entrepreneurs need first to determine the existence of the data, fact, and other forms of evidence that can be applied to promote a specific viewpoint on a matter. Second, they need to find solutions to collect this evidence to support their policy proposal (Stone, 1997).
4. A fourth key skill for policy entrepreneurs is to build compelling arguments on data and evidence they collected, and according to their knowledge of the local policy context. These arguments should be developed both to strengthen support for a policy innovation and to reduce resistance to change. When gaps in evidence and data are being used to undermine a favored position, it is the responsibility of the policy contractor to explore avenues for finding new evidence or data (Mintrom, 2019).
5. While framing information in a single way may help make it compelling to a specific audience, it is not enough to make a policy change. To ensure the success of their policy proposal, policy entrepreneurs need to engage with a large variety of stakeholders in other ways to understand their views, hopes and concerns, and share their views and perspective in a variety of fashions. This way, they can ensure a higher level of support

among a broader range of stakeholders with different interests and views on the matter at stake (Riker, 1986; Shepsle, 2003).

6. One of the main objectives of policy entrepreneurs is to challenge the status quo by bringing about policy innovations. But those who gain from the status quo seldom respond positively to those who intend to change it. Hence, it is essential for policy entrepreneurs to develop negotiation skills in order to deal with future conflict and resistance. Through negotiation, the policy entrepreneur can gain the support of a larger range of stakeholders for their policy alternative, as well as minimize the occurrence and scope of conflict and resistance from those who are against the change proposed in the policy (Mintrom, 2019).
7. The actors who succeed in bringing change in particular contexts have generally gathered relevant information from multiple sources, as early research in organizational innovation (Mohr, 1969) and spread of policy innovations (Walker, 1969) already showed. What is more, they increase substantially their probability of success when they engage in cross jurisdiction policy networks (Kammerer & Namhata, 2018; True & Mintrom, 2001). In fact, the policy-making process is often pictured as a form of continuous conversations among stakeholders (Kingdon, 1984; Majone, 1989; Mintrom, 2003). Hence, networking is a key skill for policy entrepreneurs who need to build a strong understanding of the policy networks and communities that are active around them and identify the most effective ways to engage with them (Goyal, Howlett, & Chindarkar, 2020; Mintrom, 2003).

As discussed, civil society can indeed contribute to the policy process through different tactics and according to specific skills, which enable it to influence specific policy spaces. The following section will discuss one specific type of policy entrepreneurs: civil society.

## CIVIL SOCIETY AND CITIZEN PARTICIPATION

Literature remains unclear about who policy entrepreneurs are and what their influence is on all the steps of the policy-making process (Goyal, Howlett, & Chindarkar, 2020). Policy entrepreneurs range from individuals (Cairney, 2018) to organizations and collective movement (Bakir & Jarvis, 2017; Botterill, 2013; Kinsella, NicGhabhann, & Ryan, 2017; Meijerink & Huitema, 2010; Mintrom, Salisbury, & Luetjens, 2014; Miskel & Song, 2004; Smith & Cumming, 2017). Each actor will adopt different strategies to influence the policy space and contribute to the policy-making process. This book focuses on civil society and its policy entrepreneurial role. Agenda setting is indeed often associated with one type of actor, namely civil society (Teisman & van

Buuren, 2013), since it offers formal and informal participation opportunities for civil society and social movements to express their views and influence the policy-making process.

Civil society has been a point of reference for philosophers since antiquity. Originally, civil society was defined in contrast to the state of nature. It then was defined in contrast to the state. The concept of civil society was debated along with questions around a good society, the rights and duties of citizens, the practice of politics, and collective life (Edwards, 2014). The emergence of the concept of civil society as understood today was linked to the centralization of political power in a given territory and the formation of states (Kaldor, 2004).

Civil society corresponds to "the realm of organized social life that is voluntary, self-generating, (largely) self-supporting, autonomous from the state, and bound by a legal order or set of shared rules" (Diamond, 1994, p.5). This section will first discuss the emergence of civil society and its role in nation-states, then the decline in conventional forms of citizen participation, and lastly non-conventional forms of citizen participation (i.e. global civil society, social movements).

## CIVIL SOCIETY AND NATION-STATE BUILDING

The nation-state system was born in the 17th century with the treaty of Westphalia of 1648. This treaty ended the thirty-year war between Protestants and Catholics and declared state sovereignty as the defining principle of international relations. In doing so, it put an end to the previous feudal system. Political authority became centralized in the secular state, which possessed the monopoly on the legitimate use of force (Camilleri, & Falk, 1992) and sovereignty over the land it controlled (Bodin, 1992).

A nation is defined as "a human group conscious of forming a community, sharing a common culture, attached to a clearly demarcated territory, having a common past and a common project for the future and claiming the right to rule itself" (Guibernau, 1996, p.47). A nation-state is indeed built on a territory with geographical boundaries, but also on a common consciousness of what the nation is, what nations have in common, and what differentiates them from other nations. Education, media, political events such as elections, cultural rituals and artifacts, as well as history, contribute to reaffirming and illustrating this common representation and self-awareness of a nation. This representation is grounded in a common identity based on common history, language, culture, and ethnicity.

National print-languages functioned as social integrators and contributed to the emergence of "imagined communities" (Anderson, 2006). The invention of the printing press led new entrepreneurs to print books and other material in

the vernacular language to maximize distribution and the number of readers. It enabled populations speaking local dialects to understand each other and form a common discourse. Based on this new capacity of communicating and sharing similar content, these "national print-languages" enabled the formation of "imagined communities," which led to the emergence of the first European nation-states, as Anderson (2006) contended.

The emergence of these "imagined communities" further led to repudiation of what and who was foreign: devaluation of other nations; exclusion of national, ethnic, and religious minorities, in particular Jews. But self-consciousness also provided the cultural common ground to enable individual subjects to become "citizens," and the emergence of solidarity bonds between them (Schulze, 1994). When groups of individuals share a sense of nationality, rulers can benefit from a deeper sense of social belonging (Schwarzenberger, 1941).

> Only a national consciousness, crystallized around the notion of a common ancestry, language, and history, only the consciousness of belonging to "the same" people, makes subjects into citizens of a single political community – into members who can feel responsible for one another. The nation or the Volksgeist (the unique spirit of the people – the first truly modern form of collective identity) provided the cultural basis for the constitutional state. (Habermas, 1998, p.113)

This modern understanding of "nation" led to our contemporary definition of citizenship. Hence, citizenship is legally rooted in civil rights, but also in a culturally defined community.

This new abstract form of social integration (i.e. citizenship) provided a new form of legitimation for the newly created nation-state (Habermas, 1998). Indeed, since the political authority was left without any religious foundation, the secular state had to find a source of legitimation. With this understanding, civil society was clearly distinguished from the state: it described a self-regulated group of associations that needed to be protected from the state. The value of civil society was in its role to protect pluralism, nurture constructive social norms, and as a defense against the domination of any particular group. In that sense, civil society was the foundation of a stable democratic polity (Edwards, 2014).

For Wheatley (2010), civil society has four key characteristics, namely separation from the state and private capital, self-organization, deliberation, and civility. By being separate from the state and private capital, civil society communicates the demands of the population to the state through advocacy and lobbying initiatives. Thus, civil society takes the role of representing parts of the population and putting pressure on decision-makers to adopt a decision favorable to those they represent.

As Kaldor (2004) asserts, civil society is "the medium through which one or many social contracts between individuals, both women and men, and

the political and economic centers of power are negotiated and reproduced" (pp.44–45). It encompasses a large array of organized and non-organized civilian forms of participation. In other words, it is composed of individual citizens who vote and demonstrate, as well as formally organized entities. As Rousseau ([1762] 2018) contends, citizens are both the addressees and the authors of the law in a democracy. The political system is legitimated by the free-will participation of citizens.

Civil society is constituted of "networks, norms and trust, that facilitate coordination and cooperation for mutual benefit" (Putnam, 1993, p.37). It constitutes a space for uncoerced human association (Walzer, 1995). According to Rousseau ([1762] 2018) and Wollstonecraft (1794), who saw participation as a learning process, civil society is a learning environment. As Jenkins and Carpentier (2013) remind us, civil society, "where people learn participation by participating, and where through the process of participation citizens become better citizens, is absolutely crucial. Participation allows for the performance of democracy, which is deemed an important component of the social in itself" (Jenkins & Carpentier, 2013, p.274).

As discussed, the role of civil society in society has evolved over time along with the evolution of political systems. New social norms and practices of civil society have appeared as nation-states and various forms of democracy emerged. More recently, however, many scholars and experts have raised concern about the decline of conventional participation; fewer citizens seem interested in taking part in conventional forms of citizen participation, as discussed below.

## Citizen Participation in Decline?

Liberal democracies are based on the assumption that citizens can and will take part in the governance of the commons (Parvin, 2015, 2018) through elections, referendums, and votes. Even if the objective of inclusive participation with complete political integration of the working class was never fully achieved in liberal democracies (Bobbio, 1984), universal suffrage still enabled a large part of the population to participate in the design of public policies, and ended the conflict between labor and capital interests (Jörke, 2016). A consequence of this inclusion is the adoption of citizens' fundamental liberal rights and the generalization of social welfare systems. This led to a high degree of contentment with representative democracy (Marshall, 1950; Crouch, 2004).

The conventional forms of citizen participation are in decline in many liberal democracies. Levels of citizen participation have progressively declined for about two decades (Jörke, 2016). This disengagement is particularly well illustrated by a reluctance to become an active member of political parties and an abstention from voting (Pharr & Putnam, 2000; Dalton, 2004). But not only.

Indeed, if participation levels have decreased, they also have aligned with income levels (Bartels, 2016; Gilens & Page, 2014; Solt, 2008). It has become evident that the decrease in participation corresponds mainly to the withdrawal of the most socially disadvantaged parts of the population (Birch, 2009; Schäfer, 2011, 2013) and is strongly related to inequalities in economic status and education level (Lijphart, 1997). Hence, there is a social bias in the decline of political participation: it is primarily the less educated and low-income citizens who vote less often and show less interest in politics (Jörke, 2016). Consequently, this political disengagement leads to "a weakening of the political importance of ordinary working people" (Crouch, 2004, p.29).

In his influential book entitled *Bowling Alone*, Robert Putnam (2000) concludes:

> [d]eclining electoral participation is merely the most visible symptom of a broader disengagement from community life. Like a fever, electoral abstention is even more important as a sign of deeper trouble in the body politic than as a malady itself. It is not just from the voting booth that Americans are increasingly AWOL. (p.35)

Putnam (2000) is probably one of the most well-known contributors to the academic debate revolving around the disengagement of citizens in post-industrial society, illustrated by decreasing levels of civic engagement and electoral turnout, and higher distrust toward public institutions, political leaders, and parties (Skocpol & Fiorina, 1999).

The incentives to participate in democratic processes are strongly related to inequalities in economic status and education level. The distrust in democratic institutions, and the lack of change in political processes has led to an increasingly large gap between populations and their representatives, who are perceived as "out of touch with the real world" and too far removed from the "normal" citizens' life-worlds (Jörke, 2016), which in turn, feeds this vicious circle of reduced citizen participation. This distrust of political figures can also be associated with a rejection of democracy as a whole (Hay, 2007) since it no longer succeeds in engaging and consequently representing all parts of society. In addition, the unequal distribution of civic skills, education, and access to information and technology – closely associated with inequalities in economic status and education level – contributes to the unequal decline in conventional forms of citizen participation.

The withdrawal of one part of society is preoccupying for multiple reasons. First, due to the rise of economic inequalities in the world, this less privileged part of the population is growing and could ultimately represent the majority in some countries. Second, this disengagement goes against the foundational principle of political equality (Christiano, 1996; Dahl, 2008). This disregard for hard-won participation opportunities has some tangible impacts. In many

pluralist democracies today, the results of elections are more contested than ever, elected officials are viewed less and less as representative, and social movements emerge to bypass traditional democratic processes.

However, political disengagement may be, in fact, only one side of the coin, and it is probably too premature to assume the decline of civic engagement (Norris, 2002, pp.5–7; Stolle, Hooghe, & Micheletti, 2005). The low rates of political participation refer to a minimalist definition of political participation, which does not include all the new variations and actors that constitute civil society today. This led Whiteley (2012) to ask whether it was time to update the definition of political participation. The following section will discuss these non-conventional forms of citizen participation.

**The Third Sector and New Social Movements**

At a time when conventional modes of political participation are in decline in some liberal democracies (Parvin, 2018), alternative forms of participation have emerged to express populations' demands for greater solidarity, as well as to denounce the inaction of politicians toward global issues (e.g. climate change). In the last decades, a number of causes triggered national and sometimes global social movements: environmental and social impacts of globalization, climate change, military operations, social and economic inequalities, among others. Hence, if we include social movements such as the "FridaysforFuture" youth movement to denounce the inaction of politicians toward climate change, political participation is not in decline but has changed form. "Los indignados," "Les Gilets Jaunes," or "Occupy Wall Street" are some well-known examples of such leaderless movements that aim to give a voice to those who no longer believe in the legitimacy of representative models of democracy. As discussed in a subsequent chapter, social media platforms have become a favored space for citizen expression and empowered social movements to coordinate their actions and reach out to a larger audience.

This leads Hay (2007) to conclude about citizen participation: "those with the most restrictive and conventional conceptions of political participation identify a strong and consistent pattern of declining political participation and engagement over time, whilst those with a more inclusive conception discern instead a change in the mode of political participation" (p.23). This also adopts a broader definition of citizen participation that includes both conventional (e.g. elections) and non-conventional forms of participation (e.g. demonstrations).

As Dalton (2008) argues about the decline of conventional participation in liberal democracies, "the trends in political activity represent changes in

the style of political action, and not just changes in the level of participation" (p.94). Rosanvallon and Goldhammer (2008) claim that:

> (…) democratic activity now extends well beyond the framework of electoral-representative institutions (...) The resulting system is complex but, in its own way, coherent. What these various counter-democratic powers have in common is that they describe a new architecture of separated powers and a much more subtle political dynamic than one ordinarily finds in political theory. (Rosanvallon & Goldhammer, 2008, p.249)

However, Rosanvallon and Goldhammer also argue that counter-democracy has been transformed "into a banal form of opposition" (p.190), which operates mainly in the negative, against new arguments, policy proposition, and other political orientations.

The "third sector" or the "nonprofit sector" encompasses all types of associations and movements between family and state, where membership and activities are voluntary. It includes a large variety of organizations such as NGOs, labor unions, political parties and churches, professional and business associations, community and self-help groups, and independent media. It corresponds to "a complex and dynamic ensemble of legally protected non-governmental institutions that tend to be nonviolent, self-organizing, self-reflexive, and permanently in tension, both with each other and with the governmental institutions that 'frame', constrict and enable their activities" (Keane, 2009, p.461).

Globalization and advances in communication technologies have led to an internationalization of civil society (Kaldor, Anheier, & Glasius, 2003). The advances in communication technologies, in particular internet and mobile phones, have indeed helped new social movements and associations to grow rapidly beyond the boundaries of states (Powell, 2007). For Rosenau (2003), world politics evolved and split into two: inter-state relations on one side, and on the other side various non-governmental actors who are independent of the state-centric world.

The global civil society includes a wide variety of actors with sometimes conflicting objectives: formal representative organizations such as parties, churches, lobbies, or trade unions cohabit with informal functional organizations such as charities, universities, think tanks, mass media, and with more informal social and political entities and their networks such as social forums, ad hoc activist coalitions, diasporas, networks, causes, or internationally coordinated social movements (Kaldor, Anheier, & Glasius, 2003).

As was the case within the nation-state, two important elements differentiate the global civil society conglomerate of actors from other ones: its voluntary nature (nonprofit organizations as opposed to multinational enterprises) and civility (as opposed to terrorist groups who resort to violence for accomplishing their goals) (Kaldor, Anheier, & Glasius, 2003). In addition, the "global"

element of this concept refers to various types of entities: truly global associations such as the World Wildlife Fund (WWF) with audiences all over the world (Clark, Friedman, & Hochstetler, 1998); organizations with activities or audiences in various countries but which are not really global (Florini, 2000); and local entities targeting global institutions in response to global issues (Gaventa, 2001).

Social movements from the 1970s and 1980s dealt with questions of human rights, peace, women, environment, and third world solidarity. They stem from 1968 student revolutions with new cosmopolitan values of peace and world collaboration in Western countries. These movements incarnate a new vision of the world and the first mass consideration of global issues that need to be taken care of at the global level. The 1990s is the decade that saw an unprecedented increase in the number of NGOs in the world; this slowed down afterwards thanks to "(…) political opportunities in a broadened political space, institutional weakness of the state and transnational regimes, and easier and less costly communication" (Kaldor, Moore, & Selchow, 2012, p.19). A large number of NGOs, think tanks, and scientific and professional networks were created at that period of time, in particular in the global north.

From the 1990s onwards, new social movements emerged to confront states' authority and defend the victims of globalization. These new forms of protests are cosmopolitan, modular, and autonomous (Kaldor, 2004). They are cosmopolitan, for people have become aware of a wider community rather than only the people they know. They are modular, for people can learn from others and understand their demands through new forms of communication. Finally, they are autonomous since any individual can sign a petition or write a message on a blog, a forum, or a Facebook page.

Although there is no real consensus on what constitutes social movement, Opp (2009) identifies some common features, including to form a group action and to advance a certain political or social agenda. To do so, a social movement can carry out, resist, or undo a social change. According to Glasberg and Shannon (2010), it corresponds to "organizational structures and strategies that may empower oppressed populations to mount effective challenges and resist the more powerful and advantaged elites" (p.150).

As discussed, non-conventional forms of citizen participation have emerged in the last several decades. This multitude of participatory formats reflect new social norms and practices associated with political participation. As Jenkins and Carpentier (2013) argue, they show new ways to structure the social and support a participatory democratic culture. In other words, these new social norms and practices often strengthen the motivation of individuals to care about the community and their role as citizens, but also their willingness to make a difference (Kligler-Vilenchik, McVeigh-Schultz, Weitbrecht, & Tokuhama, 2012).

## CONCLUDING REMARKS

As discussed in this chapter, the citizen–government relation is more than the delivery of governmental services. It is also about including civil society in the policy-making cycle (OECD, 2001). The Multiple Streams Framework (MSF) is a powerful conceptualization of the policy process, and specifically agenda setting (Kingdon, 1984). It argues that policy entrepreneurs (e.g. civil society) need resources (e.g. technology) and specific skills (e.g. engaging multiple audience) to develop and implement tactics (e.g. narrative reframing) through problem, policy and politics streams, to identify and exploit successfully open policy windows. Touraine (1992) contends that there cannot be any form of democracy without freedom of political choice. As Parry, Moyser, and Day (1992) contend, citizen participation corresponds to all these "action[s] by citizens which [are] aimed at influencing decisions which are, in most cases, ultimately taken by public representatives and officials" (p.16). If conventional forms of participation are in decline in some liberal democracies (Parvin, 2015, 2018), other forms of participation have developed including street protests and boycotts, leading some scholars to argue in favor of a transformation of citizen participation rather than a decline.

To strengthen citizen–government relations and citizen participation in policy making, the OECD (2001) recommends governments using digital technology for three types of actions: (1) enhancing access to information so that citizens are well informed, (2) enabling citizens to express their views on projects and societal issues that affect them in consultations, and (3) engaging citizens in decision-making processes. Information plays a crucial role throughout the policy-making process. Said differently, who provides and gains access to information, as well as who influences its distribution, gains a competitive advantage in the problem, policy and politics streams. The next chapter examines how AI affects access to and distribution of information. It also highlights who develops this technology, who benefits from it, and who is harmed by it.

## NOTE

1. Translation of: La démocratie et la liberté sont l'exception dans l'histoire de l'humanité qui est jalonnée de régimes autoritaires. Leur éclosion, leur expansion et leur avenir se confondent avec l'évolution de l'Europe et de l'Occident. (…) Démocratie, liberté, Europe sont inséparables (Sidjanski, 1979, p.13).

# REFERENCES

Aberbach, J. D. & Christensen, T. (2014). Why reforms so often disappoint. *The American Review of Public Administration*, 44(1), 3–16. https://doi.org/10.1177/0275074013504128

Ackrill, R., Kay, A., & Zahariadis, N. (2013). Ambiguity, multiple streams, and EU policy. *Journal of European Public Policy*, 20(6), 871–887.

Anderson, B. (2006). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. New York: Verso Books.

Anderson, B. & Shutes, I. (2014). *Migration and Care Labour: Theory, Policy and Politics*. Berlin: Springer.

Anderson, C. W. (1979). The place of principles in policy analysis. *American Political Science Review*, 73(3), 711–723.

Anderson, S. E., DeLeo, R. A., & Taylor, K. (2020). Policy entrepreneurs, legislators, and agenda setting: Information and influence. *Policy Studies Journal*, 48(3), 587–611.

Arklay, T., Van Acker, E., & Hollander, R. (2018). Policy entrepreneurs searching for the open-minded skeptic: A new approach to engagement in difficult policy areas. *Policy Design and Practice*, 1(2), 103–114.

Aviram, N. F., Cohen, N., & Beeri, I. (2020). Policy entrepreneurship in developing countries: A systematic review of the literature. *Public Administration and Development*, 40(1), 35–48.

Bakir, C. (2009). Policy entrepreneurship and institutional change: Multilevel governance of central banking reform. *Governance*, 22(4), 571–598.

Bakir, C. & Jarvis, D. S. L. (2017). Contextualising the context in policy entrepreneurship and institutional change. *Policy and Society*, 36(4), 465–478.

Bartels, L. M. (2016). *Unequal Democracy*. Princeton, NJ: Princeton University Press.

Baumgartner, F. R. & Jones, B. D. (1993). *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Béland, D. (2005). Ideas and social policy: An institutionalist perspective. *Social Policy & Administration*, 39(1), 1–18.

Béland, D. & Howlett, M. (2016). How solutions chase problems: Instrument constituencies in the policy process. *Governance*, 29(3), 393–409.

Berlin, I. (1969). *Four Essays on Liberty*. Oxford: Oxford University Press.

Bernhagen, P. & Marsh, M. (2007). Voting and protesting: Explaining citizen participation in old and new European democracies. *Democratisation*, 14(1), 44–72.

Birch, S. (2009). The case for compulsory voting. *Public Policy Research*, 16(1), 21–27.

Birkland, T. A. [2012] (2015). *An Introduction to the Policy Process: Theories, Concepts, and Models of Public Policy Making*. London: Routledge.

Boasson, E. L. & Huitema, D. (2017). Climate governance entrepreneurship: Emerging findings and a new research agenda. *Environment and Planning C: Politics and Space*, 35(8), 1343–1361.

Boasson, E. L. & Wettestad, J. (2014). Policy invention and entrepreneurship: Bankrolling the burying of carbon in the EU. *Global Environmental Change*, 29, 404–412.

Bobbio, N. (1984). The future of democracy. *Telos*, 1984(61), 3–16.

Bodin, J. (1992). *On Sovereignty*. Cambridge: Cambridge University Press.

Bonn, I. (2005). Improving strategic thinking: A multilevel approach. *Leadership & Organization Development Journal*, 26(5), 336–354.
Botterill, L. C. (2013). Are policy entrepreneurs really decisive in achieving policy change? Drought policy in the USA and Australia. *Australian Journal of Politics and History*, 59(1), 97–112.
Bourdieu, P. (1991). *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
Braun, M. (2009). The evolution of emissions trading in the European Union: The role of policy networks, knowledge and policy entrepreneurs. *Accounting, Organizations and Society*, 34(3–4), 469–487.
Brouwer, S. & Huitema, D. (2018). Policy entrepreneurs and strategies for change. *Regional Environmental Change*, 18(5), 1259–1272.
Cairney, P. (2018). Three habits of successful policy entrepreneurs. *Policy and Politics*, 46(2), 199–215.
Camilleri, J. A. & Falk, J. (1992). *The End of Sovereignty?: The Politics of a Shrinking and Fragmenting World*. Aldershot, UK and Brookfield, VT, USA: Edward Elgar Publishing.
Carter, N. & Jacobs, M. (2014). Explaining radical policy change: The case of climate change and energy policy under the British Labour Government 2006–10. *Public Administration*, 92(1), 125–141.
Christiano, T. (1996). Is the participation argument self-defeating? *Philosophical Studies*, 82(1), 1–12.
Christiansen, P. M. & Klitgaard, M. B. (2010). Behind the veil of vagueness: Success and failure in institutional reforms. *Journal of Public Policy*, 30(2), 183–200.
Christopoulos, D. C. (2006). Relational attributes of political entrepreneurs: A network perspective. *Journal of European Public Policy*, 13(5), 757–778.
Clark, A. M., Friedman, E. J., & Hochstetler, K. (1998). The sovereign limits of global civil society: A comparison of NGO participation in UN world conferences on the environment, human rights, and women. *World Politics*, 51(1), 1–35.
Cobb, R. W. & Elder, C. D. (1983). *Participation in American Politics: The Dynamics of Agenda-Building*. Baltimore, MD: Johns Hopkins University Press.
Cochran, C. L. & Malone, E. F. (1999). *Public Policy: Perspectives and Choices*. Sydney: McGraw-Hill College.
Copeland, P. & James, S. (2014). Policy windows, ambiguity and commission entrepreneurship: Explaining the relaunch of the European Union's economic reform agenda. *Journal of European Public Policy*, 21(1), 1–19. https://doi.org/10.1080/13501763.2013.800789
Cox, R. H. & Béland, D. (2013). Valence, policy ideas, and the rise of sustainability. *Governance*, 26(2), 307–328.
Crouch, C. (2004). *Post-Democracy*. Cambridge: Polity Press.
Crow, D. A. (2010a). Local media and experts: Sources of environmental policy initiation? *Policy Studies Journal*, 38(1), 143–164.
Crow, D. A. (2010b). Policy entrepreneurs, issue experts, and water rights policy change in Colorado. *Review of Policy Research*, 27(3), 299–315.
Cummings, C. (2015). Fostering innovation and entrepreneurialism in public sector reform. *Public Administration and Development*, 35(4), 315–328. https://doi.org/10.1002/pad.1735
Dahl, R. A. (2008). *On Political Equality*. New Haven, CT: Yale University Press.

Dalton, R. (2004). *Democratic Challenges, Democratic Choices, The Erosion of Political Support in Advanced Industrial Democracies*. Oxford: Oxford University Press.
Dalton, R. J. (2008). Citizenship norms and the expansion of political participation. *Political Studies*, 56(1), 76–98.
Danzig, R. (2018). *Technology Roulette: Managing Loss of Control as Many Militaries Pursue Technological Superiority*. Center for New American Security: Washington, DC, USA, 2018. https://www.cnas.org/ publications/reports/technology-roulette [Accessed 21 August 2021].
Diamond, L. (1994). Rethinking civil society: Toward democratic consolidation. *Journal of Democracy*, 5(1994), 4–18.
Dixit, A. K. & Nalebuff, B. (2008). *The Art of Strategy: A Game Theorist's Guide to Success in Business and Life*. New York: W. W. Norton & Company.
Easton, David (1965). *A Systems Analysis of Political Life*. New York: John Wiley.
Edwards, M. (2014). *Civil Society*. Cambridge: Polity Press.
European Union (EU). (2012). Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union. *Official Journal of the European Union*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2010:083:FULL&from=EN Publications Office of the European Union, Luxembourg, 2020 [Accessed 21 August 2021].
Florini, A. (2000). *The Third Force: The Rise of Transnational Civil Society*. Washington, DC: The Carnegie Endowment for International Peace.
Fowler, L. (2019). Problems, politics, and policy streams in policy implementation. *Governance*, 32(3), 403–420.
García-Albacete, G. M. (2011). "Continuity or generational change? A longitudinal study of young people's political participation in Western Europe" (Doctoral dissertation). Mannheims universitet.
Gamson, W. A. & Modigliani, A. (1987). The changing culture of affirmative action. In: Braungart, R. G. and Braungart, M. M. (eds.), *Research in Political Sociology*. Greenwich, CT: JAI Press, Vol. 3, pp.137–177.
Gaventa, J. (2001). *Global Citizen Action*. London: Earthscan Publications.
Gerston, L. N. (2004). *Public Policy Making: Process and Principle*. Armonk, NY: M. E. Sharpe.
Gilens, M. & Page, B. I. (2014). Testing theories of American politics: Elites, interest groups, and average citizens. *Perspectives on Politics*, 12(3), 564–581.
Glasberg, D. S. & Shannon, D. (2010). *Political Sociology: Oppression, Resistance, and the State*. London: Sage Publications.
Goldfinch, S. & Hart, P. T. (2003). Leadership and institutional reform: Engineering macroeconomic policy change in Australia. *Governance*, 16(2), 235–270.
Goyal, N., Howlett, M., & Chindarkar, N. (2020). Who coupled which stream(s)? Policy entrepreneurship and innovation in the energy–water nexus in Gujarat, India. *Public Administration and Development*, 40(1), 49–64.
Guibernau, M. (1996). *Nationalisms*. Cambridge: Polity Press.
Haas, P. M. (1992). Introduction: Epistemic communities and international policy coordination. *International Organization*, 46(1), 1–35.
Habermas, J. (1998). *Inclusion of the Other: Studies in Political Theory*. New York: John Wiley & Sons.
Hart, P. S., Chinn, S., & Soroka, S. (2020). Politicization and polarization in COVID-19 news coverage. *Science Communication*, 42(5), 679–697.
Hay, C. (2007). *Why We Hate Politics*, Vol. 5. Cambridge: Polity Press.

Hedge, D. M., Lester, J. P., & Stewart, J. (2008). *Public Policy: An Evolution Approach*. Belmont, CA: Thomson Wadsworth.
Heikkila, T., Pierce, J. J., Gallaher, S., Kagan, J., Crow, D. A., & Weible, C. M. (2014). Understanding a period of policy change: The case of hydraulic fracturing disclosure policy in Colorado. *Review of Policy Research*, 31(2), 65–87.
Herweg, N., Huß, C., & Zohlnhöfer, R. (2015). Straightening the three streams: Theorising extensions of the multiple streams framework. *European Journal of Political Research*, 54(3), 435–449.
Herweg, N., Zahariadis, N., & Zohlnhöfer, R. (2018). The multiple streams framework: Foundations, refinements, and empirical applications. In: Weible, C. M. and Sabatier, P. A. (eds.), *Theories of the Policy Process*. London: Routledge, pp.17–53.
Howlett, M. (1998). Predictable and unpredictable policy windows: Institutional and exogenous correlates of Canadian federal agenda-setting. *Canadian Journal of Political Science/Revue canadienne de science politique*, 31(3), 495–524.
Howlett, M., McConnell, A., & Perl, A. (2015). Streams and stages: Reconciling Kingdon and policy process theory. *European Journal of Political Research*, 54(3), 419–434.
Howlett, M., McConnell, A., & Perl, A. (2016). Weaving the fabric of public policies: Comparing and integrating contemporary frameworks for the study of policy processes. *Journal of Comparative Policy Analysis: Research and Practice*, 18(3), 273–289.
Howlett, M., McConnell, A., & Perl, A. (2017). Moving policy theory forward: Connecting multiple stream and advocacy coalition frameworks to policy cycle models of analysis. *Australian Journal of Public Administration*, 76(1), 65–79.
Hoyt, L. (2006). Importing ideas: The transnational transfer of urban revitalization policy. *International Journal of Public Administration*, 29(1–3), 221–243.
Huppe, P. L. & Hill, M. J. (2006). The three action levels of governance: Re-framing the policy process beyond the stages model. In: Peters, G. and Pierre, J. (eds.), *Handbook of Public Policy*. London: Sage, pp.13–30.
Hysing, E. (2009). Greening transport: Explaining urban transport policy change. *Journal of Environmental Policy and Planning*, 11(3), 243–261.
Inglehart, R. & Catterberg, G. (2002). Trends in political action: The developmental trend and the post-honeymoon decline. *International Journal of Comparative Sociology*, 43(3–5), 300–316.
Iyengar, S. (1991). *Is Anyone Responsible? How Television Frames Political Issues*. Chicago, IL: University of Chicago Press.
Jenkins, H. & Carpentier, N. (2013). Theorizing participatory intensities: A conversation about participation and politics. *Convergence*, 19(3), 265–286.
Jenkins-Smith, H. C., Nohrstedt, D., Weible, C. M., & Ingold, K. (2018). The advocacy coalition framework: An overview of the research program. In: Weible, C. M. and Sabatier, P. A. (eds.), *Theories of the Policy Process*. London: Routledge, pp.135–171.
Jones, B. D. (1977). Distributional considerations in models of government service provision. *Urban Affairs Review*, 12(3), 291–312.
Jörke, D. (2016). Political participation, social inequalities, and special veto powers. *Critical Review of International Social and Political Philosophy*, 19(3), 320–338.
Kaase, M. & Marsh, A. (1979). Political action: A theoretical perspective. In: Barnes, S. H. and Kaase, M. (eds.), *Political Action: Mass Participation in Five Western Democracies*. Beverly Hills, CA: Sage Publications, pp.27–56.

Kalafatis, S. E., Grace, A., & Gibbons, E. (2015). Making climate science accessible in Toledo: The linked boundary chain approach. *Climate Risk Management*, 9, 30–40.
Kaldor, M. (2004). *Global Civil Society: An Answer to War*. Cambridge: Polity Press.
Kaldor, M., Anheier, H., & Glasius, M. (2003). *Global Civil Society 2003*. Oxford: Oxford University Press.
Kaldor, M., Moore, H. L., & Selchow, S. (2012). *Global Civil Society 2012* (Vol. 248). Basingstoke: Palgrave Macmillan.
Kalil, T. (2017). Policy entrepreneurship at the White House: Getting things done in large organizations. *Innovations: Technology, Governance, Globalization*, 11(3–4), 4–21.
Kammerer, M. & Namhata, C. (2018). What drives the adoption of climate change mitigation policy? A dynamic network approach to policy diffusion. *Policy Sciences*, 51(4), 477–513.
Keane, K. (2009). Civil society, definitions and approaches. In: Anheier, H. K. and Toepler, S. (eds.), *International Encyclopedia of Civil Society*. Berlin: Springer Science & Business Media, pp.461–464.
Kingdon, J. (1984). *Agendas, Alternatives and Public Policies*. Boston: Little, Brown.
Kingdon, J. W. (1995). *Agendas, Alternatives, and Public Policies* (2nd edn). New York: HarperCollins College Publisher.
Kingdon, J.W. (2003). *Agendas, Alternatives, and Public Policies* (3rd edn). New York: Longman.
Kinsella, S., NicGhabhann, N., & Ryan, A. (2017). Designing policy: Collaborative policy development within the context of the European capital of culture bid process. *Cultural Trends*, 26(3), 233–248.
Kitzinger, J. (2004). *Framing Abuse: Media Influence and Public Understanding of Sexual Violence against Children*. London: Pluto Press.
Knaggård, Å. (2015). The multiple streams framework and the problem broker. *European Journal of Political Research*, 54(3), 450–465.
Knaggård, Å. (2016). Framing the problem: Knowledge brokers in the multiple streams approach. In: Zohlnhöfer, R. and Rüb, F. W. (eds.), *Decision-Making Under Ambiguity and Time Constraints: Assessing the Multiple Streams Framework*. Colchester: ECPR Press, pp.109–123.
Kolkman, D. (2020). The usefulness of algorithmic models in policy making. *Government Information Quarterly*, 37(3), 101488.
Krause, R. M. (2011). Symbolic or substantive policy? Measuring the extent of local commitment to climate protection. *Environment and Planning C, Government & Policy*, 29(1), 46–62.
Kwon, M., Jang, H. S., & Feiock, R. C. (2014). Climate protection and energy sustainability policy in California cities: What have we learned? *Journal of Urban Affairs*, 36(5), 905–924.
Lasswell, H. D. (1951). *The Policy Orientation. The Policy Sciences*. Stanford: Stanford University Press.
Lasswell, H. D. (1971). *A Preview of Policy Sciences*. Amsterdam: Elsevier Publishing Company.
Leach, M., Scoones, I., & Stirling, A. (2010). *Dynamic Sustainabilities*. London: Earthscan.
Levi-Faur, D. & Vigoda-Gadot, E. (2006). New public policy, new policy transfers: Some characteristics of a new order in the making. *International Journal of Public Administration*, 29(4–6), 247–262.

Lijphart, A. (1997). Unequal participation: Democracy's unresolved dilemma presidential address. *American Political Science Review*, 91(1), 1–14.
Lippmann, W. (1922). *Public Opinion*. New York: Harcourt Brace.
Littlejohn, S. W. & Foss, K. A. (2010). *Theories of Human Communication*. Long Grove, IL: Waveland Press.
Lovell, H. (2009). The role of individuals in policy change: The case of UK low energy housing. *Environment and Planning C, Government & Policy*, 27(3), 491–511.
Majone, G. (1989). *Evidence, Argument, and Persuasion in the Policy Process*. New Haven, CT: Yale University Press.
Mallett, A. & Cherniak, D. (2018). Views from above: Policy entrepreneurship and climate policy change on electricity in the Canadian Arctic. *Regional Environmental Change*, 18(5), 1323–1336.
Maor, M. (2017). Policy entrepreneurs in policy valuation processes: The case of the coalition for environmentally responsible economies. *Environment and Planning C: Politics and Space*, 35(8), 1401–1417.
Marshall, T. H. (1950). *Citizenship and Social Class* (Vol. 11, pp.28–29). New York: Cambridge University Press.
McCombs, M. E., Shaw, D. L., & Weaver, D. H. (2014). New directions in agenda-setting theory and research. *Mass Communication and Society*, 17(6), 781–802.
McFadgen, B. K. (2019). Connecting policy change, experimentation, and entrepreneurs: Advancing conceptual and empirical insights. *Ecology and Society*, 24(1), 30–50. https://www.jstor.org/stable/26796918 [Accessed 24 March 2022]
Meijerink, S. & Huitema, D. (2010). Policy entrepreneurs and change strategies: Lessons from sixteen case studies of water transitions around the globe. *Ecology and Society*, 15(2), 17.
Mintrom, M. (1997). Policy entrepreneurs and the diffusion of innovation. *American Journal of Political Science*, 41(3), 738–770.
Mintrom, M. (2003). *People Skills for Policy Analysts*. Chicago: Georgetown University Press.
Mintrom, M. (2019). So you want to be a policy entrepreneur? *Policy Design and Practice*, 2(4), 307–323.
Mintrom, M. & Norman, P. (2009). Policy entrepreneurship and policy change. *Policy Studies Journal*, 37(4), 649–667.
Mintrom, M. & Salisbury, C. (2014). Policy entrepreneurs, creative teamwork, and policy change. In: Alexander, D. and Lewis, J. M. (eds.), *Making Public Policy Decisions: Expertise, Skills and Experience*. London: Routledge, pp.129–145.
Mintrom, M., Salisbury, C., & Luetjens, J. (2014). Policy entrepreneurs and promotion of Australian state knowledge economies. *Australian Journal of Political Science*, 49(3), 423–438.
Mintrom, M. & Vergari, S. (1996). Advocacy coalitions, policy entrepreneurs, and policy change. *Policy Studies Journal*, 24(3), 420–434.
Mintzberg, H., Ahlstrand, B., & Lampel, J. (1998), *Strategy Safari. A Guided Tour through the Wilds of Strategic Management*. New York: The Free Press.
Miskel, C. & Song, M. (2004). Passing reading first: Prominence and processes in an elite policy network. *Educational Evaluation and Policy Analysis*, 26(2), 89–109.
Mohr, L. B. (1969). Determinants of innovation in organizations. *The American Political Science Review*, 63(1), 111–126.
Mouffe, C. (2000). *The Democratic Paradox*. New York: Verso Books.

Mukherjee, I. & Howlett, M. (2015). Who is a stream? Epistemic communities, instrument constituencies and advocacy coalitions in public policy-making. *Politics and Governance*, 3(2), 65–75.
Mukhtarov, F. & Gerlak, A. K. (2013). River Basin organizations in the global water discourse: An exploration of agency and strategy. *Global Governance*, 19(2), 307–326.
Navot, D. & Cohen, N. (2015). How policy entrepreneurs reduce corruption in Israel. *Governance*, 28(1), 61–76.
Nay, O. (2012). How do policy ideas spread among international administrations? Policy entrepreneurs and bureaucratic influence in the UN response to AIDS. *Journal of Public Policy*, 32(1), 53–76.
Nohrstedt, D. (2011). Shifting resources and venues producing policy change in contested subsystems: A case study of Swedish signals intelligence policy. *Policy Studies Journal*, 39(3), 461–484.
Norris, P. (2002). *Democratic Phoenix: Reinventing Political Activism*. Cambridge: Cambridge University Press.
Oborn, E., Barrett, M., & Exworthy, M. (2011). Policy entrepreneurship in the development of public sector strategy: The case of London health reform. *Public Administration*, 89(2), 325–344.
OECD (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publications.
Oliver, T. R. (2006). The politics of public health policy. *Annual Review of Public Health*, 27(1), 195–233.
Opp, K. D. (2009). *Theories of Political Protest and Social Movements: A Multidisciplinary Introduction, Critique, and Synthesis*. London: Routledge.
Ostrom, E. (2010). Analyzing collective action. *Agricultural Economics*, 41, 155–166.
Palmer, J. R. (2015). How do policy entrepreneurs influence policy change? Framing and boundary work in EU transport biofuels policy. *Environmental Politics*, 24(2), 270–287.
Paquet, M. (2015). Bureaucrats as immigration policy-makers: The case of subnational immigration activism in Canada, 1990–2010. *Journal of Ethnic and Migration Studies*, 41(11), 1815–1835.
Parry, G., Moyser, G., & Day, N. (1992). *Political Participation and Democracy in Britain*. Cambridge: Cambridge University Press.
Parvin, P. (2015). Is deliberative democracy feasible? Political disengagement and trust in liberal democratic states. *The Monist*, 98(4), 407–423.
Parvin, P. (2018). Democracy without participation: A new politics for a disengaged era. *Res Publica*, 24(1), 31–52.
Patterson, C., Semple, S., Wood, K., Duffy, S., & Hilton, S. (2015). A quantitative content analysis of UK newsprint coverage of proposed legislation to prohibit smoking in private vehicles carrying children. *BMC Public Health*, 15(1), 1–7.
Perry, B. & Uuk, R. (2019). AI governance and the policymaking process: Key considerations for reducing AI risk. *Big Data and Cognitive Computing*, 3(2), 26.
Peters, B. G., Jordan, A., & Tosun, J. (2017). Over-reaction and under-reaction in climate policy: An institutional analysis. *Journal of Environmental Policy and Planning*, 19(6), 612–624.
Petridou, E. (2014). Theories of the policy process: Contemporary scholarship and future directions. *Policy Studies Journal*, 42, S12–S32.
Pharr, S. J. & Putnam, R. D. (eds.) (2000). *Disaffected Democracies: What's Troubling the Trilateral Countries?* Princeton, NJ: Princeton University Press.

Popper, K. R. (1959). The propensity interpretation of probability. *The British Journal for the Philosophy of Science*, 10(37), 25–42.
Powell, F. W. (2007). *The Politics of Civil Society: Neoliberalism or Social Left?* London: Policy Press.
Prateek, G., Kumar, K., Kar, P., & Krishnan, A. (2021). Civil society as policy entrepreneur in agriculture and forestry sectors amidst COVID-19 lockdown in India. *Journal of Asian Public Policy*, 1–23.
Putnam, R. (1993). The prosperous community: Social capital and public life. *The American Prospect*, 4(13), 35–42.
Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.
Quintelier, E. & Hooghe, M. (2013). The impact of socio-economic status on political participation. In: Demetriou, K. N. (ed.), *Democracy in Transition: Political Participation in the European Union.* Berlin, Heidelberg: Springer, pp.273–289.
Rabe, B. G. (2004). *Statehouse and Greenhouse: The Emerging Politics of American Climate Change Policy*. Washington, DC: Brookings Institution Press.
Riker, William H. (1986). *The Art of Political Manipulation*. New Haven: Yale University Press.
Roberts, N. C. & King, P. J. (1991). Policy entrepreneurs: Their activity structure and function in the policy process. *Journal of Public Administration Research and Theory*, 1(2), 147–175.
Rosanvallon, P. & Goldhammer, A. (2008). *Counter-Democracy: Politics in an Age of Distrust* (Vol. 7). Cambridge: Cambridge University Press.
Rosenau, J. N. (2003). *Distant Proximities: Dynamics beyond Globalization*. Princeton, NJ: Princeton University Press.
Rousseau, J. J. ([1762] 2018). *The Social Contract and Other Later Political Writings*. Cambridge: Cambridge University Press.
Sabatier, P. A. (1988). An advocacy coalition framework of policy change and the role of policy oriented learning therein. *Policy Sciences*, 21(2/3), 129–168.
Sabatier, P. A. (1999). *Theories of the Policy Process*. Oxford: Westview Press.
Sabel, C. F. & Zeitlin, J. (2008). Learning from difference: The new architecture of experimentalist governance in the EU. *European Law Journal*, 14(3), 271–327.
Sabel, C. F. & Zeitlin, J. (2012). Experimentalist governance. In: Levi-Faur, D. (ed.), *The Oxford Handbook of Governance*. Oxford: Oxford University Press, pp.169–187.
Schäfer, A. (2011). Republican liberty and compulsory voting. MPIfG Discussion Paper, No. 11/17. Cologne, Germany: Max Planck Institute for the Study of Societies (MPIfG).
Schäfer, A. (2013). Liberalization, inequality and democracy's discontent. In: Streeck, W. and Schäfer, A. (eds.), *Politics in the Age of Austerity*. Cambridge: Polity Press, pp.169–195.
Schattschneider, E. E. (1960). *Party Government*: *American Government in Action*. Piscataway, NJ: Transaction Publishers.
Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49(4), 103–122.
Schulze, H. (1994). *Staat und Nation in der Europäischen Geschichte*. Munich: C.H. Beck.
Schwarzenberger, G. (1941). *Power Politics: Introduction to the Study of International Relations and Postwar Planning*. London: Jonathan Cape.

Shepsle, K. A. (2003). Losers in politics (and how they sometimes become winners): William Riker's heresthetic. *Perspective on Politics*, 1(2), 307–315.
Shpaizman, I., Swed, O., & Pedahzur, A. (2016). Policy change inch by inch: Policy entrepreneurs in the Holy Basin of Jerusalem. *Public Administration*, 94(4), 1042–1058.
Sidjanski, D. (1979). *Europe Élections de la démocratie européenne*. Paris: Stanké.
Sidney, M. S. (2017). Policy formulation: Design and tools. In: Fischer, F. and Miller, G. J. (eds.), *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. London: Routledge, pp.105–114.
Skocpol, T. & Fiorina, M. P. (1999). Making sense of the civic engagement debate. In: Skocpol, T. and Fiorina, M. P. (eds.), *Civic Engagement in American Democracy*. Washington, DC: Brookings Institution Press, pp.1–23.
Smith, V. & Cumming, J. (2017). Implementing pay for performance in primary health care: The role of institutional entrepreneurs. *Policy and Society*, 36(4), 523–538.
Solt, F. (2008). Economic inequality and democratic political engagement. *American Journal of Political Science*, 52(1), 48–60.
Stewart, J., Jr., Hedge, D. M., & Lester, J. P. (2007). *Public Policy: An Evolutionary Approach*. Toronto: Nelson Education.
Stolle, D., Hooghe, M., & Micheletti, M. (2005). Politics in the supermarket: Political consumerism as a form of political participation. *International Political Science Review*, 26(3), 245–269.
Stone, D. A. (1997). *Policy Paradox: The Art of Political Decision Making*. New York: W.W. Norton.
Teisman, G. R. & van Buuren, A. (2013). Models for research into decision-making processess. In Araral, E. (ed.), *Routledge Handbook of Public Policy*. London: Routledge, pp.299–320.
Touraine, A. (1992). What is democracy? *UNESCO Courier*. https://en.unesco.org/courier/novembre-1992/what-democracy [Accessed 21 August 2021].
True, J. & Mintrom, M. (2001). Transnational networks and policy diffusion: The case of gender mainstreaming. *International Studies Quarterly*, 45(1), 27–57.
Tuchman, G. (1978). *Making News: A Study in the Construction of Reality*. New York: Free Press.
Ugur, M. & Yankaya, D. (2008). Policy entrepreneurship, policy opportunism, and EU conditionality: The AKP and TÜSĐAD experience in Turkey. *Governance*, 21(4), 581–601.
United Nations (2001). *Universal Declaration of Human Rights*. https://www.un.org/en/universal-declaration-human-rights/ [Accessed 21 August 2021].
Verba, S. (1967). Democratic participation. *The Annals of the American Academy of Political and Social Science*, 373(1), 53–78.
Verba, S. & Nie, N. H. (1972). *Participation in America: Social Equality and Political Democracy*. New York: Harper& Row.
Verba, S., Schlozman, K. L., & Brady, H. E. (1995). *Voice and Equality: Civic Voluntarism in American Politics*. Cambridge, MA: Harvard University Press.
Verger, A. (2012). Framing and selling global education policy: The promotion of public–private partnerships for education in low income contexts. *Journal of Education Policy*, 27(1), 109–130.
Voß, J. P. & Simons, A. (2014). Instrument constituencies and the supply side of policy innovation: The social life of emissions trading. *Environmental Politics*, 23(5), 735–754.

Walker, J. L. (1969). The diffusion of innovations among the American states. *American Political Science Review*, 63(3), 880–899.
Walzer, M. (1995). *Toward a Global Civil Society* (Vol. 1). New York: Berghahn Books.
Weick, K. E. (1983). Managerial thought in the context of action. In: Srivastva, S. (ed.), *The Executive Mind. New Insights on Managerial Thought and Action*. San Francisco, CA: Jossey-Bass, pp.221–242.
Wheatley, J. (2010). Civil society in the Caucasus: Myth and reality. *Caucasus Analytical Digest*, 12/10, Zurich: CSS.
Whiteley, P. (2012). *Political Participation in Britain: The Decline and Revival of Civic Culture*. Basingstoke: Palgrave Macmillan.
Wikström, M., Eriksson, L., & Hansson, L. (2016). Introducing plug-in electric vehicles in public authorities. *Research in Transportation Business and Management*, 18, 29–37.
Wollstonecraft, M. (1794). *Historical and Moral View of the Origin and Progress of the French Revolution and the Effect it Has Produced in Europe* (Vol. 1). Chicago, IL: Johnson.
Wolmer, W., Keeley, J., Leach, M., Mehta, L., Scoones, I., & Waldman, L. (2006). *Understanding Policy Processes: A Review of IDS Research on the Environment*. Brighton: Institute of Development Studies.
Zahariadis, N. (2003). *Ambiguity and Choice in Public Policy: Political Decision Making in Modern Democracies*. Washington, DC: Georgetown University Press.
Zahariadis, N. (2007). The multiple streams framework: Structure, limitations, prospects. In: Sabatier, P. (ed.), *Theories of the Policy Process*. Boulder, CO: Westview Press, pp.65–92.
Zahariadis, N. & Exadaktylos, T. (2016). Policies that succeed and programs that fail: Ambiguity, conflict, and crisis in Greek higher education. *Policy Studies Journal*, 44(1), 59–82. https://doi.org/ 10.1111/psj.12129
Zhu, X. (2008). Strategy of Chinese policy entrepreneurs in the third sector: Challenges of technical infeasibility. *Policy Sciences*, 41(4), 315–334.

# 3. AI and information dissemination: Challenging citizens' access to relevant and reliable information

## INTRODUCTION

> Designing a personalized ranking system for more than 2 billion people (all with different interests) and a plethora of content to select from presents significant, complex challenges. This is something we tackle every day with News Feed ranking. Without machine learning (ML), people's News Feeds could be flooded with content they don't find as relevant or interesting, including overly promotional content or content from acquaintances who post frequently, which can bury the content from the people they're closest to. Ranking exists to help solve these problems, but how can you build a system that presents so many different types of content in a way that's personally relevant to billions of people around the world? We use ML to predict which content will matter most to each person to support a more engaging and positive experience. Models for meaningful interactions and quality content are powered by state-of-the-art ML.
>
> (Lada, Wang, & Yan, 2021)

As more and more citizens are connected to the internet through their smartphones and other web-browsing devices, information is disseminated quickly and widely. In the European Union, about nine out of ten netizens use search engine websites at a minimum once a week, and six out of ten use an online social media platform at least once a week (European Union, 2021). In a digital age, individuals, organizations, and governments have access to a wide array of communication and communication technologies (ICTs), channels, and techniques to produce and share information. As Sidjanski (2000) argues, "[w]e are moving from a logic of energy, exclusive and leading to centralized hierarchical systems, to a logic of information based, like biological systems, on complementarity, synergy, and interdependence" (p.203). The generalization of ICTs, and in particular online platforms, triggers the emergence of new patterns of interactions among these actors, based on the values that Don Tapscott identified in 2008 as transparency, participation, and collaboration.

The European Commission acknowledges that online platforms play a key role in today's social and economic life by enabling European citizens to

access information online, and for businesses to benefit from e-commerce advantages. Although internet penetration rates are not the same throughout Europe, the majority of the EU population is connected to the internet and uses social media platforms on a regular basis. Most social media platforms only require an email address and an internet connection to join (Fuchs, 2012). In other words, they are accessible and offer a large array of tools to interact, network, and debate. Hence, social media platforms are potentially the ideal context for people "to challenge discourses, share alternative perspectives and publish their own opinions" (Loader & Mercea, 2011, p.759).

The generalization of social media platforms and the rapid adoption of smartphones in the European Union (EU) are indicative of the high levels of internet penetration and almost constant connectivity of citizens. These platforms were created to foster dialogue among citizens from diverse backgrounds and origin, or as Facebook puts it "[to] give people the power to build community and bring the world closer together" (Facebook, n.d.). However, online platforms have not only offered a new creative space for self-expression and participatory communication (Jenkins, 2006a), but they have also redefined communication (Langlois & Elmer, 2013).

In 2018, it was estimated that 56% of individuals living in the European Union (EU) take part in social media, and 48% used social media platforms every day or nearly every day in 2019 (Statistica, 2020). According to the Global Web Index, in 2020 internet users aged 16 to 64 spent an average of 2h24m daily on social media on one device or another. In terms of number of active users in the world in January 2020, Facebook (2.4 billion) still dominates, followed by YouTube (2bn), WhatsApp (1.6bn), FB Messenger (1.3bn), WeChat (1.1bn), and Instagram (1bn) (We Are Social, 2020). We cannot fail to notice that the Facebook group owns four out of the six top social media platforms with the most users in the world (Facebook, WhatsApp, FB Messenger, and Instagram). As Nieborg (2015) argues:

> All platforms are equal, but some are more equal than others. Facebook's capabilities to leverage network effects are infinitely bigger than any other platform currently up and running in the social media universe. (Van Dijck & Poell, 2015, p.4)

The concept of online platform[1] describes a broad range of applications. The EU Commission argues that "[o]nline platforms share key characteristics including the use of information and communication technologies to facilitate interactions (including commercial transactions) between users, collection and use of data about these interactions, and network effects which make the use of the platforms with most users most valuable to other users" (European Commission, n.d.). Artificial intelligence, and more precisely machine learning algorithms (MLA) are the core piece of the functioning of online platforms,

and allows them to generate wealth and provide their services to individuals and organizations. For instance, Facebook's newsfeed algorithm selects the most relevant content for each user to show in priority in their newsfeed (Lada, Wang, & Yan, 2021).

Algorithms are used to govern many aspects of our society and economy (Janssen & Kuk, 2016). As Osoba and Welser IV (2017) argue, an algorithm can be defined as "a computable function that can be implemented on computer systems. Machine learning algorithms can also update their behavior in response to experience (input data) and performance metrics" (Osoba & Welser IV, 2017 cited in Hoorens & Lupiáñez-Villanueva, 2019, p.21). They can be described as automatized decision processes, or step-by-step instructions to process inputs into outputs (Stone, 1971). Today, most algorithms consist of an aggregate of numerous algorithms that function as a computer program (Sandvig, 2014). Some of these algorithms, including the ones considered in this book, are powered by AI, which means in simple terms that they have the capacity to learn from data and adapt their code accordingly. This is what we refer to as MLA. And MLAs are at the core of the success of social media platforms. It is indeed not possible to study social media platforms without considering MLAs.

The OECD (2001) recommends that governments use digital technology to strengthen the citizen–government relations through three types of actions: (1) enhancing access to information so that citizens are well informed, (2) enabling citizens to express their views on projects and societal issues that affect them in consultations, (3) engaging citizens in decision-making processes. These three types of action must also be designed and implemented according to the principles of equity and inclusion, in order to avoid any discrimination within the population, and between the actors involved in the policy-making process. This chapter focuses on the first type of action recommended by the OECD (2001) to strengthen citizen–government relations: ensure that information is "complete, objective, reliable, relevant, easy to find and to understand" (p.11) for citizens. Since a large number of citizens access information through online platforms, this chapter explores how AI affects information distribution on online platforms. On social media platforms, information distribution is indeed specific. AI and more precisely MLAs enable social media platforms to automate information distribution flows. In these digital spaces, citizens adopt specific behavior and information is diffused differently from other media.

Online platforms have had an impact on policy making (Chun & Luna Reyes, 2012) as they provide a new channel that facilitates social networking, crowdsourcing, and interactive dialogues between citizens and other stakeholders in the policy-making process. This chapter first briefly discusses the conceptual challenges to define social media platforms. Then it examines the specific characteristics of information flows on social media platforms, and the

role of AI in selecting, filtering, ranking, and diffusing information. Finally, it examines advocacy efforts from civil society organizations and social movements to creative advocacy tactics and strategies on social media platforms to have their voice heard.

## A CONCEPTUAL CHALLENGE

The development of the internet enabled many other technologies to emerge, including the World Wide Web and social media platforms. Web 1.0 and web 2.0 are often used to describe two eras: one where technology only enabled content to be produced and posted on a website for others to "read only"; the latter which allows the audience to interact, respond, comment, and even produce new content. Web 2.0 is based on the new capacity offered to users to self-generate content online and to interact with other users without the interference of elite media and traditional sources of authority. The previous version of the web only allowed one-way communication; the online user was a passive consumer of information. With web 2.0, online users have become both consumers and producers of information.

Social media belongs to the second era, but is rooted in the first. Therefore the separation between the two is not as clear-cut as sometimes presented: it is more a continuum "from one that prioritizes a social imagination of indefinite strangers, to one that vacillates between imagined strangers and numerable, identifiable, individuals" (Ankerson, 2015, p.11).

Social media platforms present many challenges as an object of study, among which is their conceptualization. This is not without consequences. Indeed, for citizens and political leaders to reach an opinion about a new technology or an issue, they must be able to clearly identify it, delineate its scope, and assess its positive and negative consequences.

Social media platforms emerged at the beginning of 2000. From 2005 until 2010, most studies concentrated on the user and the new creative space offered. The following five years focused on the professional use of such platforms by public and private organizations, as well as social movements (Van Dijck & Poell, 2015), to promote products and interact with their audiences in order to raise awareness and advertise for products, and coordinate actions.

To study social media platforms, the researcher faces four challenges. The first one is the conceptualization challenge of social media platforms, which is difficult due to both the perpetual transformation of the social media landscape and to the opacity of the activities of some large corporate social media platforms such as Facebook (Obar & Wildman, 2015). Social media platforms are indeed perpetually evolving, making it difficult to identify precise boundaries around the concept. These platforms include a large range of computer- and smartphone-based applications, with specific cultural and national features,

that are continuously being launched, relaunched and abandoned in different countries in the world. Moreover, other technologies also provide the same service as social media platforms – connecting people in the world; in other words, should we consider the phone, the fax machine or email as social media platforms?

To respond to this conceptualization challenge, Obar and Wildman (2015) identified some key characteristics in the literature that can help distinguish social media platforms from other information and communication technologies (ICTs).[2]

- First, social media platforms are web 2.0 internet-based applications. In the 1990s, the services provided through the internet were mainly reading on the World Wide Web (Web) and consuming audio and video clips on commercial media. The change occurred with the emergence of the social web, also called web 2.0. This technological advance and ideology change (Kaplan & Haenlein, 2010) contributed to enabling users to also interact with others and the content they produced. Hence, users are what Ritzer and Jurgenson (2010) call prosumers. This is why web 2.0 is a place where content can be "continuously modified by all users in a participatory and collaborative fashion" (Kaplan & Haenlein, 2010, p.61).
- Second, user-generated content is the lifeblood of social media platforms. The content produced and shared by users, whether a blog entry, a photo on Instagram, a high score on Candy Crush, or a "like" on Facebook, is the backbone of social media platforms. Without this content, social media platforms would actually become a ghost town.
- Third, profiling is the backbone of social media platforms (Boyd & Ellison, 2008). Although the forms of identification differ from one social media platform to another, they either require users to create a user profile in order to use their service, or they create a profile in their database.[3] This profiling allows social media platforms to connect users with each other, and offer services that users expect such as comparing between gaming scores, sharing voting results, liking content, etc.
- Fourth, social media platforms connect users with each other either by creating a list of individuals to connect with (Facebook and Snapchat call them friends, Twitter and Instagram call them followers, and LinkedIn connections) or through location-based or content preference (such as Yik Yak). "The nature and nomenclature of these connections may vary from site to site" (Boyd & Ellison, 2008, p.211).

However, even with these four criteria, it remains a challenge to define the limits of what social media is today.

Social media platforms can be categorized according to their geographical scope and utility. On the one hand, "universal" social media platforms, such as Facebook, provide users with a new digital space on a global stage where they can either create personal profiles or an official page for an organization. This type of platform allows for personal but also for professional interactions. Intimate conversations can take place, as well as business transactions. Users write comments, share views, post and watch videos, play games, respond to quizzes, and so on. On the other hand, some specialized social media platforms cater either to dedicated geographical parts of the world population (e.g. WeChat in China and some parts of Asia) or certain professions (e.g. LinkedIn for professionals or ResearchGate and Academia for academics), while others offer tools for specific activities, such as publishing videos (e.g. YouTube), photos (e.g. Instagram), text content (e.g. Blogger) among others.

The second challenge to studying social media platforms is the opacity of most activities of social media platforms. As Langlois and Elmer (2013) argue, social media platforms may well look like a transparent platform where a large array of communication acts take place. However, their transparency is limited to these communication acts; in fact, this is the only aspect of all their activities that is visible. Actually, content production and networking are only the tip of the iceberg: the business model of Facebook for instance is to collect as much data as possible from its users. In this context, Facebook not only records content produced and distributed on its platforms, but it also collects a large spectrum of metadata, including:

> [S]pecific information about the profile of the user sending out a message, the users receiving that message, about how users interact with a message by reading or not reading it, "liking" it, sharing it (…) time lapses, time spent on a page or scrolling, pauses in the communication process, silences that might seem non-communicational but that still yield information as to what a user is reading or deciding not to react to, as well as previous communication acts that give a specific communication act a discursive and social context (…) content users access and interact with at different times of the day and night and in different social settings (at work, home, or with friends), but also of how users themselves act on different platforms and how they share content across a multitude of platforms. (Langlois & Elmer, 2013, pp.2–3)

Through the big data collected, Facebook intends to "enhance, format, encode and diagnose communication" (p.4) with the purpose of not only promoting for-profit content, such as advertising, but "to tap into everyday life in order to try and refashion it from the inside" (p.4).

The third challenge is ontological, meaning that researchers need to focus their attention away from what is being said to how it is being processed and rendered. As Langlois and Elmer (2013) contend, "We must expand from the

study of communication as signs or discourse, to include the study of communication as data collection, storage, and processing" (pp.2–3).

Finally, the fourth challenge is methodological: how to analyze the different layers of data collected and produced on and by corporate social media platforms, including content that is not available to the researcher. The answer that Langlois and Elmer (2013) propose is through the concept of digital object: "Digital objects, as previously explained, are the elements that compose social media platforms in specific context: a 'like' button is a digital object, for instance, as is a comment or any other kinds of text" (p.11).

The digital object counts three layers or characteristics. First, it is a media object: digital objects are constituted of content and form. Second, it a network object: digital objects connect informational networks: "'Liking' a news story usually means that other hidden informational networks are activated: profiling networks, for instance, that will then adapt the content of the ads on a news website to the Facebook profile of the user" (Langlois & Elmer, 2013, p.11). Third, it is a phatic object (Miller, 2008): it is an action of presence; it positions users within their network, and establishes position and relation within this network of users and digital objects.

As discussed in this section, social media platforms present many challenges as an object of study. This definition challenge is not without consequences. Indeed, for citizens and political leaders to have an opinion about a new technology or an issue, they must be able to clearly identify it, delineate its scope, and assess its positive and negative consequences. This is particularly challenging when the concept is vague and encompasses a large variety of applications.

## AI AND ACCESS TO INFORMATION

Over the years, and more recently with the UK referendum Brexit, concerns associated with the rapid adoption of social media platforms, and the transformation of their business models into advertising giants and data brokers, led the general public and policy makers to change their perception of social media platforms and question their role in society. For instance, in 2012, a popular TED video of academic Eli Pariser (2012) called upon the experts and tech companies of Silicon Valley to adapt their services and products so that citizens could have access to pluralist sources of information and opinions. More recently, *The Guardian* unveiled the Cambridge Analytica scandal (Cadwalladr & Graham-Harrison, 2018). This section examines key characteristics of information distribution on social media platforms.

## Information Overload

On social media platforms, users can both consume and produce information (Fuchs, 2012). Citizens contribute for the most part to the content published online by recording videos and taking photos that are posted on social media (Goldkind, 2015). User-generated content greatly increases the availability of information. Emotional messages and images attract the attention of users on social media (Stieglitz & Dang-Xuan, 2013), and can increase the momentum for a specific issue. This is particularly true when health is threatened by the issue at stake, and when messages are illustrated by tragic images of individuals suffering. A news item will tend to spread across numerous online clusters and networks, with comments added to original framing. It may create momentum, a buzz, and become viral. In that case, it provides a space for many actors to intervene and position themselves. Social media channels offer indeed a space for citizens, consumers, businesses, politicians, and experts to have a say and make their voices heard. In most cases, public figures have an account on Twitter or Facebook, and are often publicly addressed on social media.

The liking, commenting, and sharing on social media trigger a rapid dissemination of news items. Users on social media spread information by reacting to a news item, thus contributing to the emergence of a momentum. This self-reinforcing process is strengthened by the networking nature of social media platforms and how their MLA are designed: MLAs of social media platforms tend to favor viral content, meaning content that triggers an emotional reaction, whether because it is sensationalist or extreme. The distancing offered by online tools, and the feeling – sometimes true – of anonymity, do not favor profound analytical exchange of views: it can be "difficult to keep online conversations from devolving into either name-calling or blather" (Shirky, 2008, p.50).

The feedback loops between a news item and the responses of social media users lead to an exaggeration of reality, often associated with related media hype (Vasterman, Yzermans, & Dirkzwager, 2005) and virality. These momenta represent a new prospect for various social actors to influence the general public and policy makers. The dynamics associated with the global character of the environment and social media platforms imply that local news can quickly become a global issue.

Moreover, this quantitative increase does not necessarily translate into a qualitative increase of information. As Jenkins argues, authors who claim that social media platforms contribute to access to information "make no claims on objectivity; they are often unapologetically partisan; they deal often with rumors and innuendos; and as we will see, there is some evidence they

are mostly read by people who already agree with their author's stated views" (Jenkins, 2006b, p.216).

**Filtering Content**

Consequently, citizens and organizations face an overload of information, and there is a necessity to organize and prioritize the information produced. Recommender systems were first developed in the early-to-mid 1990s to address the information overload by building prediction models that estimate how much the user will like each of a large set of items (Konstan & Riedl, 2012). They were progressively developed in public and private settings, including universities and e-commerce websites. Their replacement are the MLAs, which select and rank content for users online. They help internet users overcome the overload of information online, and make decisions in terms of what to read, listen to, or watch.

Since each individual has unique tastes and interests, the information selection must be done at the individual level. This is where algorithms enter into play in information diffusion: their role is to select the most relevant information for each netizen at any particular time. This selection is done through the data collected about each individual in the past, and cross-analyzed with the data about others. This data collection and analysis enable the MLA to categorize each individual user.

However, this categorization is not without issues. For instance, some scholars revealed mismatched face recognition with a racial bias according to the Fitzpatrick skin tones. Wilson, Hoffman, and Morgenstern (2019) conducted a study "on recent examples of ML and vision systems displaying higher error rates for certain demographic groups than others" (p.1). Also, Amazon's facial recognition technologies were criticized for mismatching members of Congress in 2018 (Snow, 2018). Moreover, while errors made either by human beings while developing algorithms or by algorithms themselves emphasize racism or sexism, there are "several cases that demonstrate how racism and sexism are part of the architecture and language of technology" (Noble, 2018, p.9). Even if "Google's algorithms have admittedly changed, such that a search for 'black girls' does not yield nearly as many pornographic results now as it did in 2011. Nonetheless, new instances of racism and sexism keep appearing in news and social media" (Noble, 2018, p.10). Noble (2018) demonstrates how algorithms are not only biased but can cause harm to gender or ethnic groups.

Moreover, the information selection done by MLAs contributes to keeping individual users in what Eli Pariser (2012) calls filter bubbles. He described the role of MLAs in filtering content accessible to netizens and users of social media platforms. The filter bubble phenomenon highlights the fact that MLAs

of social media platforms tend to restrict their users' access to information based on what the user already "liked," consequently not contributing to the plurality of sources. MLAs use data collected through behavioral tracking and cookies technologies to identify what content is the most relevant for the user at any given time. Their objective is to keep individuals online as long as possible in order to collect personal data and then target them with ads. This is problematic in pluralist democracy, since it does not allow citizens to be well informed, meaning to be exposed to relevant and pluralist sources of information. Eli Pariser gives the example in 2012 of Facebook deleting the comments of his friends who were from the other side of the political spectrum.

Since MLA are designed and managed by private companies, the criteria – at least one of them – of such algorithms is to increase the profitability of the company. It increases its revenue through collecting data from citizens in order to target advertising. This means that from the tech company's perspective, citizens need to remain online as long as possible. Consequently, the algorithm will provide information that will not necessarily push the citizen to think differently, but rather it will provide easy-to-digest and entertaining information.

This aspect of information dissemination is now well known and documented. As Devaux (2019) contends, algorithms allow large tech companies such as Facebook and Instagram to continuously adapt bespoke content to netizens, positioning these MLA as de facto gatekeepers to information.

## A Plurality of Gatekeepers

MLA have become an additional gatekeeper in the information ecosystem of European pluralist democracies: they determine what is newsworthy, rank content according to pre-defined criteria, and filter the access to content. In that context, MLAs of social media platforms play a substantial role in democracy and can have an impact on citizen participation (although it remains difficult to assess it). However, MLAs are not the only gatekeepers today. The press remains a strong gatekeeper, as well as influencers on social media platforms.

The role of gatekeeper also has a direct impact on the press. As Tandoc, Jenkins, and Kraft (2019) argue, "while platforms have provided another channel for publishers to disseminate their content, platforms have also taken audience attention away from traditional news sites" (p.675). Indeed, mainstream media had to adapt to the production and distribution model of content for the digital and social media environment. More precisely, they had to adapt to the algorithms of social media platforms and web search engines to regain the revenue they lost when their audiences shifted from offline paid news to online free news. In many European countries, citizens access news through social media platforms and web search engines, which place the algorithms in the role of gatekeepers.

Mainstream media outlets have adapted their 24-hour cycle of news production to the "online first model" (Vergeer, 2018, p.38). The online first model means that citizens must produce a piece of news faster than their competitors in order to bring revenue to their media outlet. So, journalists had to adapt to this new environment as well; they also had to produce not only reliable information, but fast information (Vergeer, 2018). In other words, journalists need to compete with sensationalist infotainment content that is easy to consume and share – at least so more than analytical or in-depth content. Consequently, revenue generated from advertising disappears from well-researched outlets and flows into sensationalist entertainment or infotainment outlets. In other words, the "online first model," which is linked to how algorithms were developed, favors information that is quickly produced and quickly online. The current digital environment does not favor well-researched content, which requires time before being published.

On social media platforms, some influencers tend to dominate the online conversations: Barzilai-Nahon (2008) defines this new form of filtering information "networked gatekeeping." In other words, some celebrities, famous journalists and bloggers, political leaders and well-known entrepreneurs, act as gatekeepers on social networks (Shaw, 2012), filtering communication flows from top to bottom (Castells, 2013, p.71). For instance, an abundance of research showed the impact of some opinion leaders on the virality of content on social media platforms (Nahon & Hemsley, 2013; Wu & Wang, 2011).

As discussed, social media platforms have developed specific tools to automatically select the information most relevant to each user. This is not without consequences for pluralist democracy, since it is based on the assumption that citizens have access to pluralist sources of information. Recently, policy makers, including from the European Union, called upon social media platforms to improve their algorithms accordingly, and some launched campaigns to raise awareness among the population about the critical perspective they need to adopt when consulting information online.

## Echo Chambers

Echo chambers are a well-known and well-established phenomenon that appeared on social media platforms, where ideas and beliefs are reinforced by communication and repetition within a homogenous group of users: "What we now know about both links and individual behavior supports the general view that many people are mostly hearing more and louder echoes of their own voices" (Sunstein, 2006, p.55). It is closely associated to the concept of homophily, which describes the fact that we tend to associate with others who are like us (Jenkins, 2006b) and "always move elsewhere if the group reaches

conclusions that run counter to [our] own beliefs or desires" (Jenkins, 2006b, p.231). This happens online and in particular on social media platforms.

Echo chambers and homophily have an impact on the plurality of views accessible to users on social media platforms. Since individuals associate with others who are similar to them, false information can spread very quickly through an individual's network. There is no one to challenge a piece of news that is shared among peers: one message is reiterated as an echo of one opinion and reinforced over and over again in one community, and without any counter-argumentation (Jamieson & Cappella, 2008). Hence, access to information is limited in terms of points of view and diversity. Moreover, political debate on social media platforms is often reduced to superficial exchanges. As Halpern and Gibbs contend, "most [social media users] are not debating rationally or deeply in this media. This suggests that political exchanges on social media may be more superficial in nature, rather than being characterized by in-depth debate or deliberation, and calls into question their efficacy" (Halpern & Gibbs, 2013, p.1166). In fact, users of social media platforms tend to leave the discussion before any meaningful exchange between different points of view can take place (Kruse, Norris, & Flinchum, 2018).

Hence, the echo chamber phenomena contributes to the quick dissemination and acceptance of false news. MLA of social media platforms have proven to be as effective for distributing bespoke content as they are for directing advertising and false information. Thanks to the low cost of entry, access to a large audience, and limited accountability, vast disinformation campaigns have occurred in recent years.

## SOCIAL MEDIA PLATFORMS AND ONLINE ADVOCACY

The great diversity of social media presents a difficulty to conceptualize this phenomenon. However, social media have allowed the emergence of a new space of expression for civil society in liberal democracies. It should be noted at this point that social media have also become a space for surveillance and censorship in some countries. The chapter on surveillance discusses this topic in more detail. In addition, social media have also become a power issue where large-scale disinformation campaigns are launched by political groups and governments. The chapter on disinformation discusses this topic in more detail.

Nevertheless, social media has become a favored place for civil society to express itself. In particular, they allow (1) to reach a wider audience more quickly and (2) to coordinate their actions at a lower cost and without the need for centralized administration.

ICTs reduce indeed the cost of communication, while increasing its speed and outreach. Thanks to the internet and mobile technologies, information is accessible to almost everyone on the planet. Internet and mobile penetration do not cease to increase. New models of organizations emerge with more open, horizontal, and dynamic structures. Civil society organizations have now the possibility to communicate on a global scale, which can strengthen their power (Edwards, 2014). Given their low cost, ICTs were rapidly adopted by the global civil society to coordinate its activities, access information, take part in global debates, raise funds, acquire new members, and organize international events.

Disruptive technologies such as the internet and social media gave birth to the knowledge society and the Net Generation: born with the third screen (Castells, 1996), the Net Generation or digital grown-ups to quote Don Tapscott, do not use mobile phones to call, but rather to tweet, take a photo, record a video and share content. These new generations know more about the dominant technology than their parents (Tapscott, 2008). They are used to produce content and actively customize the information they wish to obtain through Twitter, RSS news feeds, news agency websites, blogs, and Facebook pages. They need unique, tailor-made and real-time solutions that respond directly to their needs and desires. The success of online streaming music and videos is a good example of this change: instead of watching television, they choose to stream their favorite series online when and where they want. Location and time become irrelevant.

Sometimes, the Net Generation becomes more visible: the Arab spring has shown how the use of new ICTs can support the organization of massive street protests. New ICTs provide opportunities for people to be politically active. What trigger these protests are not new ICTs, but rather injustice, lack of jobs, repression, violence and economic disparity. Tools used to fight this revolution are, however, no longer lethal weapons: new ICTs allow the Net Generation to raise awareness, denounce Human Rights violations and call for help. Internet and mobile phones are used by the youth to spread the word, decide on a meeting point, join forces on an issue and influence the society they are living in. These newly created networks give to the population a real sense of participation.

These digital grown-ups represent a large part of society: 80 million in the USA alone, compared to 78 million of baby boomers (Tapscott, 2008). In other parts of the world such as Asia and Africa, they represent an even larger part of society and become a powerful force of change. Public figures have become aware of this new opportunity to be in touch with the general public. President Obama, for instance, created an online platform for his two presidential campaigns, where citizens could discuss issues and make their views available to

the President. This openness and transparency made him very popular among young voters.

This generation with ownership of online tools becomes a powerful force of change: "thus, the industrial society, by educating citizens and by gradually organizing the economy around knowledge and information, prepared the ground for the empowering of the human mind when new information technologies became available" (Castells, 1996, p.31). Similarly to the printing press invention that enabled Renaissance thinkers to share knowledge and planted the seed for the European technological dominance a few centuries later, the generalization of new ICTs contributed to the emergence of a new type of society with new patterns of interaction, namely transparency, collaboration, and participation (Tapscott, 2008) described as follows.

Because they were gradually and massively adopted by all actors, and in particular individuals, social media platforms are composed of multiple online spaces, where traditional stakeholders such as states and international organizations (IOs) interact and compete with non-state actors, including individuals, civil society organizations, and businesses at the local, national, and international levels. It was estimated that in 2020 nearly 60% of the world population will use social media (We Are Social, 2020). A wide array of nonprofit actors uses social media for advocacy purposes, enabling them to set the international agenda, to organize group actions and collaborate on a global scale, and finally to collect information for advocacy purposes, whether marketing or public relations.

Social media platforms can be characterized by their ability to enable many-to-many communication on a global scale. They can be described as a collection of instruments. Previously, interpersonal communication was only possible on dedicated media such as telephones. Similarly, traditional broadcasting media such as TV, radio, and newspapers, delivered information to a large number of people, but this was one-way communication only. Social media is an innovation first in the sense that it blends one-to-one with one-to-many communication streams, but in addition it enables individuals to become producers of information and not only consumers, and allows broadcast media to target its messages individually. Second, social media is an innovation because it allows two-way communication, where organizations and individuals interact on a permanent basis on social networking platforms such as Facebook, Twitter, and LinkedIn. Relationships are at the core of social media platforms (Goldkind, 2015).

Among those who benefited most from the generalization of ICTs and more precisely from computerization, internet penetration, and the rapid adoption of mobile phones, are the individuals and CSOs that gained new capacity to reach out to a larger range of stakeholders (Lovejoy & Saxton, 2012), such as donors, volunteers, online and offline media, and the general public. The diversity of

the nonprofit community provides an excellent opportunity to examine the use of social media, which differs from one organization to another. Civil society actors have developed a capacity for outreach in order to gain new donors, members, and volunteers. In a knowledge society, information becomes both a resource and a force that transforms decisional and non-decisional processes. Many CSOs have developed web 2.0 public relations campaigns to raise awareness about specific issues, which has triggered the emergence of new online spaces where legitimacy is built through the inclusion of a wider number of stakeholders, starting with the general public. Citizens become actors on the international stage by interacting with CSOs, IOs, states, and businesses on social media. In that context, social media platforms can change how civil society actors relate to each other, as well as the dynamics of their interaction.

This gradual technological shift has also dramatically transformed the media landscape. For example, traditional media, such as printed matter, had to reinvent themselves, adding an online component to the paper version. Most newspapers, if not all of them, have a website, social media channels, and specific content developed for their online audiences. Radio adapted some of its content to podcast formats, and TV channels developed apps and online streaming platforms. Some citizens have become journalists and started reporting on events from the field. Social media, and in particular Twitter, has become the place to be when it comes to finding or publishing the latest information. The most recent news items are no longer found in newspapers but online. In a large number of countries, citizens read the news predominantly on social media.

This rapid transformation of the media landscape became of high interest for researchers and policy makers, since this change affects the role and power of media. In particular, a considerable number of studies analyze the use of social media by CSOs (Roback, 2013), and most argue that CSOs still need to develop additional capacity in order to make optimal use of social media platforms (Bortree & Seltzer, 2009).

The wide variety of social media channels facilitates the exchange of user-generated text, audio and video files, and instruments, and have empowered individuals and organizations to develop and conduct advocacy campaigns (Guo & Saxton, 2014). Social media platforms rely mostly on the content generated by organizations (Tredinnick, 2006) to attract potential donors, members, volunteers, petition signatories, and digital ambassadors who can convey their message to their own personal networks.

According to the Oxford online dictionary, advocacy is to publicly support or recommend a particular cause or policy. Lovejoy and Saxton (2012, p.341) identified three key communicative tactics used by nonprofits for their advocacy work on social media: one-way information, community building and call

to action. First, information is when the organization presents itself, its activities, future events and provides information that is relevant to its audience. Second, community is when the organization interacts with its audience and aims to develop a community. Lastly, action is when the organization sends out a call for action such as to participate in an event, donate, or share a type of media. This is not the only way to categorize social media tactics, but one that corresponds to a general understanding of the three main types of actions found on social media.

Guo and Saxton (2014, p.71) divide these tactics into three similar categories, but with different names. First, reaching out to people corresponds to the information function in the previous model. Advocacy work aims to educate and raise awareness about specific issues. Second, keeping the flame alive corresponds to community and dialogue. It allows organizations to deepen existing relations with their audiences and develop new ties with others. The organization's aim here is to nurture its audience and build an active community of supporters. Third, stepping up to action corresponds to the organization calling for action and asking its supporters to mobilize. What is interesting to see is the convergence between the two analyses performed on social media, which indicates that these three categories do in fact represent how CSOs manage their social media platforms.

CSOs have increasingly adopted technologically intensive media to influence various stakeholders (FitzGerald & McNutt, 1999); among the most commonly used are social media platforms. Indeed, individuals are more easily approached through these new instruments for a number of reasons. First, social media users can see at the same time and in the same place a mix of personal contact updates and sponsored content that promotes a product, a service, or a cause. This means that the line between personal content and sponsored content is blurred, which can greatly benefit advocacy campaigns in terms of credibility and trust. One naturally has more trust in information distributed in a secure and intimate space than in the outside world. Second, individuals regularly visit social media channels. As mentioned previously, European citizens have massively adopted social media platforms for their personal and professional occupations. This means that the audience on social media platforms is vast, which makes it a valuable space for promotion.

Third, social media proposes a large number of promotional instruments, ranging from ads with highly detailed and targeted segmentation tools, to contests and customized applications. Social media allow CSOs to rapidly and efficiently identify their target audiences, organizations with a common agenda, and empathic individuals. For advocacy professionals, this is a gold mine, since they can segment their audiences at a level unforeseen so far. Fourth, each element of a campaign, whether an application, a message, a post,

or an ad, can be designed for a specific part of the population, and can further be monitored and readjusted depending on the success of the campaign.

What is more, CSOs also use social media platforms to coordinate their actions (e.g. street protests, stunt, boycotts) at a low cost and without heavy and centralized administrative processes. Since their early developments, social media platforms have become a favored space for social movements and facilitated their emergence (although did not trigger them). "Los indignados," "Les Gilets Jaunes" or "Occupy Wall Street" are some well-known examples of such leaderless movements that aim to give a voice to civil society. Social movements use social media platforms to coordinate their actions and take to the streets to express their concern about global issues, such as the immobility of states in the face of the climate issue. For instance, the grassroots movement "Right To Know Rally" began its activities with one page on Facebook (Adamoli, 2012) and rapidly grew into an international movement spread over 400 cities in North America and Europe. More recently, the well-known movements of "Extinction Rebellion" and "Fridaysforfuture" have made extensive use of social media platforms since the creation of the two movements to reach out and coordinate their actions with very little governance.

Through its multiple platforms, social media allows personal interactions, peer recognition, and the strengthening of group norms (Valenzuela, 2013), which in turn stimulate individual and community identity construction, two crucial components of political conduct. The various aspects linked to the environment, such as use of land, food safety, and ecosystem management among others, are crucial components of identity construction, and have led to the emergence of numerous online communities. Furthermore, individual forms of online protests are increasingly associated with lifestyle elements, which results in the personalization of global issues (Bennett & Segerberg, 2011). This implies that in the war of narratives on social media platforms between multinational corporations, governments, and the global civil society, the framing of narratives becomes central and influences the definition of global public issues.

As illustrated in this section, social media platforms enable civil society to reach out to a larger audience through online advocacy campaigns, and coordinate their actions nationally and globally. Although it is true that social media platforms allow civil society to raise awareness and citizens to interact and access information, the picture is incomplete. As a result, over the last couple of years, the narrative associated with social media platforms has evolved to include new concerns as discussed in the next section. MLAs are at the core of these concerns.

## CONCLUDING REMARKS

As discussed in this chapter, AI enable these platforms to automate information distribution flows, and in particular to rank, filter, and diffuse information. This leads to phenomena such as filter bubble and echo chambers, which does not support citizens' efforts to access "complete, objective, reliable, relevant, easy to find and to understand" information. In this context, the algorithms of social media platforms (in their current development stage) do not benefit civil society and its capacity to make well-informed decisions. From this perspective, AI does not the strengthen the citizen–government relation.

However, social media platforms also offer civil society organizations and social movements an unprecedented opportunity to develop creative advocacy campaigns in order to have their voice heard. They offer a new avenue for civil society to influence the policy-making process. Leach, Stirling, and Scoones (2010) coined the term policy space to describe these new contexts where stakeholders such as civil society can "influence policy formulation and implementation" by "informal (media writings, campaigns, social media advocacy) means." From that perspective, AI strengthens the citizen–government relation.

## NOTES

1. We will use the terms social media and online platforms as synonyms.
2. ICTs comprise all technologies that help gather, distribute, produce, consume, and store information, including print and broadcast media, channels of communication (satellite, cable), telecommunications (phone, web), computers, and storage devices. Definition from Singh (2002, p.2).
3. For instance, Yik Yak does not require users to provide a real name or real photo when they sign up. However, as indicated in the Yik Yak privacy policy, the platform creates a unique user profile for each individual in their database, and they track the geolocation of the data produced, the mobile device used, comments and the vote inputs. This profiling allows Yik Yak to deliver the functionality that users expect such as location-based messages, profile scores, message scores, etc.

## REFERENCES

Adamoli, G. (2012). Social media and social movements: A critical analysis of audience's use of Facebook to advocate food activism offline. *Electronic Theses, Treatises and Dissertations*.

Ankerson, M. S. (2015). Social media and the "read-only" web: Reconfiguring social logics and historical boundaries. *Social Media+ Society*, 1(2), 1–12.

Barzilai-Nahon, K. (2008). Toward a theory of network gatekeeping: A framework for exploring information control. *Journal of the American Society for Information Science and Technology*, 59(9), 1493–1512.

Bennett, W. L. & Segerberg, A. (2011). Digital media and the personalization of collective action. *Information, Communication & Society*, 14, 197–215.
Bortree, D. & Seltzer, T. (2009). Dialogic strategies and outcomes: An analysis of environmental advocacy groups' Facebook profiles. *Public Relations Review*, 35, 317–319.
Boyd, D. M. & Ellison, N. B. (2008). Social network sites: Definition, history and scholarship. *Journal of Computer-Mediated Communication*, 13, 210–230.
Cadwalladr, C. & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. *The Guardian*, 17, 22.
Castells, M. (1996, second edition, 2009). *The Rise of the Network Society, The Information Age: Economy, Society and Culture Vol. I*. Malden, MA and Oxford: Blackwell.
Castells, M. (2013). *Communication Power*. Oxford: Oxford University Press.
Chun, S. A. & Luna-Reyes, L. F. (2012). Social media in government. *Government Information Quarterly*, 29(4), 441–445.
Devaux, F. (2019, August). The true processing in memory accelerator. In *2019 IEEE Hot Chips 31 Symposium (HCS)*. IEEE Computer Society, pp.1–24.
Edwards, Michael (2014). *Civil Society*. Cambridge: Polity Press.
European Commission (n.d.). Online platforms. Glossary. Luxembourg: Publications Office of the European Union. https://ec.europa.eu/digital-single-market/en/glossary#Online_Platforms [Accessed 21 August 2021].
European Union (2021). Digital economy and society statistics – households and individuals. Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals#Internet_usage [Accessed 29 March 2022].
Facebook (n.d.). What is Facebook's mission statement? https://investor.fb.com/resources/default.aspx [Accessed 21 August 2021].
FitzGerald, E. & McNutt, J. G. (1999). Electronic advocacy in policy practice: A framework for teaching technologically based practice. *Journal of Social Work Education*, 35(3), 331–341.
Fuchs, C. (2012). The political economy of privacy on Facebook. *Television and New Media*, 13(2), 139–159.
Goldkind, L. (2015). Social media and social service: Are nonprofits plugged in to the digital age? *Human Service Organizations: Management, Leadership & Governance*, 39(4), 380–396.
Guo, C. & Saxton, G. (2014). Tweeting social change: How social media are changing nonprofit advocacy. *Nonprofit and Voluntary Sector Quarterly*, 41(1), 57–79.
Halpern, D. & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168.
Hoorens, S. & Lupiáñez-Villanueva, F. (eds.) (2019). *Study on Media Literacy and Online Empowerment Issues Raised by Algorithm-Driven Media Services*. Luxembourg: Publications Office of the European Union.
Jamieson, K. H. & Cappella, J. N. (2008). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment*. Oxford: Oxford University Press.
Janssen, M. & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371–377.
Jenkins, H. (2006a). *Fans, Gamers, and Bloggers: Exploring Participatory Culture*. New York: New York University Press.
Jenkins, H. (2006b). *Convergence Culture*. New York: New York University Press.

Kaplan, A. M. & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of social media. *Business Horizons*, 53(1), 59–68.
Konstan, J. A. & Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1–2), 101–123.
Kruse, L. M., Norris, D. R., & Flinchum, J. R. (2018). Social media as a public sphere? Politics on social media. *The Sociological Quarterly*, 59(1), 62–84.
Lada, A., Wang, M., & Yan, T. (2021). How machine learning powers Facebook's News Feed ranking algorithm. *Facebook Engineering*. https://engineering.fb.com/2021/01/26/ml-applications/news-feed-ranking/ [Accessed 21 August 2021].
Langlois, G. & Elmer, G. (2013). The research politics of social media platforms. *Culture Machine*, 14, 1–17.
Leach, M., Stirling, A. C., & Scoones, I. (2010). *Dynamic Sustainabilities: Technology, Environment, Social Justice*. London: Routledge.
Loader, B. D. & Mercea, D. (2011). Networking democracy? Social media innovations and participatory politics. *Information, Communication and Society*, 14(6), 757–769.
Lovejoy, K. & Saxton, G. D. (2012). Information, community, and action: How nonprofit organizations use social media. *Journal of Computer-Mediated Communication*, 17(3), 337–353.
Miller, V. (2008). New media, networking and phatic culture. *Convergence*, 14(4), 387–400.
Nahon, K. & Hemsley, J. (2013). *Going Viral*. Cambridge: Polity.
Nieborg, D. B. (2015). Crushing candy: The free-to-play game in its connective commodity form. *Social Media+ Society*, 1(2).
Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: New York University Press.
Obar, J. A. & Wildman, S. (2015). Social media definition and the governance challenge: An introduction to the special issue. *Telecommunications Policy*, 39(9), 745–750.
OECD (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publications.
Osoba, O. A. & Welser IV, W. (2017). *An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence*. Santa Monica, CA: Rand Corporation.
Oxford Dictionary (n.d.). Definition of advocacy. https://en.oxforddictionaries.com/definition/advocacy [Accessed 21 August 2021].
Pariser, Eli. (2012). Beware of online filter bubbles. *TED*. https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles? [Accessed 21 August 2021].
Ritzer, G. & Jurgenson, N. (2010). Production, consumption, prosumption: The nature of capitalism in the age of the digital "prosumer". *Journal of Consumer Culture*, 10(1), 13–36.
Roback, A. J. (2013). Uncovering motives for social networking site use among practitioners at non-profit organizations. *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*. New York: ACM.
Sandvig, C. (2014). Seeing the sort: The aesthetic and industrial defense of "the algorithm". *Journal of the New Media Caucus*. http://median.newmediacaucus.org/art-infrastructures-information/seeing-the-sort-the-aesthetic-and-industrial-defense-of-the-algorithm/ [Accessed 29 March 2022].
Shaw, A. (2012). Centralized and decentralized gatekeeping in an open online collective. *Politics & Society*, 40(3), 349–388.
Shirky, C. (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. London: Penguin Books.

Sidjanski, D. (2000). *The Federal Future of Europe: From the European Community to the European Union*. Ann Arbor, MI: University of Michigan Press.
Singh, J. P. (2002). Information technologies and the changing scope of global power and governance. In: Rosenau, J. N. (ed.), *Information Technologies and Global Politics: The Changing Scope of Power and Governance*. Albany: State University of New York Press, pp.1–38.
Snow, J. (2018). Amazon's face recognition falsely matched 28 members of congress with mugshots. ACLU blog. https://www.aclu.org/blog/privacy-technology/surveillance-technologies/amazons-face-recognition-falsely-matched-28 [Accessed 21 August 2021].
Statistica (2020). *Social Media in Europe*. https://www.statista.com/topics/4106/social-media-usage-in-europe/ [Accessed 21 August 2021].
Stieglitz, S. & Dang-Xuan, L. (2013). Emotions and information diffusion in social media – sentiment of microblogs and sharing behavior. *Journal of Management of Information Systems*, 29, 217–248.
Stone, H. S. (1971). *Introduction to Computer Organization and Data Structures*. New York: McGraw-Hill.
Sunstein, C. (2006). *Republic 2.0*. Princeton, NJ: Princeton University Press.
Tandoc Jr, E. C., Jenkins, J., & Craft, S. (2019). Fake news as a critical incident in journalism. *Journalism Practice*, 13(6), 673–689.
Tapscott, D. (2008). *Wikinomics: How Mass Collaboration Changes Everything*. New York: Portfolio, Penguin Group.
Tredinnick, L. (2006). Web 2.0 and business: A pointer to the intranets of the future? *Business Information Review*, 23(4), 228–234.
Valenzuela, S. (2013). Unpacking the use of social media for protest behavior: The roles of information opinion expression, and activism. *American Behavioral Scientist*, 57, 920–942.
Van Dijck, J. & Poell, T. (2015). Social media and the transformation of public space. *Social Media+ Society*, 1(2), 2056305115622482.
Vasterman, P., Yzermans, C. J., & Dirkzwager, A. J. (2005). The role of the media and media hypes in the aftermath of disasters. *Epidemiologic Reviews*, 27(1), 107–114.
Vergeer, M. (2018). Incorrect, fake, and false. Journalists' perceived online source credibility and verification behavior. *Observatorio*, 12(1), 37–52.
We Are Social. (2020). *Digital 2020: 3.8 billion people use social media*. https://wearesocial.com/blog/2020/01/digital-2020-3-8-billion-people-use-social-media [Accessed 21 August 2021].
Wilson, B., Hoffman, J., & Morgenstern, J. (2019). Predictive inequity in object detection. *arXiv preprint arXiv:1902.11097*.
Wu, H. L. & Wang, J. W. (2011). An empirical study of flow experiences in social network sites. *PACIS 2011 Proceedings*, 215. https://aisel.aisnet.org/pacis2011/215 [Accessed 29 March 2022].

# 4. AI in public and private forms of surveillance: Challenging trust in the citizen–government relations

## 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

time. 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

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

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

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

> 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 (Bashyakarla, Hankey, Macintyre, 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,
- New Zealand's Government Communications Security Bureau.

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 geolocalization 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

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 surveillance 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 provider 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 government 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 Axciom, 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. Qrioulsy 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

Acton, R. (2018). The hyper-personalised future of political campaigning. *CAPX*. https://capx.co/the-hyper-personalised-future-of-political-campaigning/ [Accessed 21 August 2021].

Acxiom Corporation. (2017). *Acxiom US Products Privacy Policy*. https://www.acxiom.com/wp-content/uploads/2017/03/ US-Products-Privacy-Policy_072516.pdf [Accessed 21 August 2021].

Adarsh, M. (2019). How Beacon technology is reshaping election campaign marketing. *Beaconstac*. https://blog.beaconstac.com/2018/04/how-beacon-technology-is-reshaping-election-campaign-marketing/ [Accessed 21 August 2021].

Agnew, G. (2003). Developing a metadata strategy. *Cataloging & Classification Quarterly*, 36(3–4), 31–46.
Aho, B. & Duffield, R. (2020). Beyond surveillance capitalism: Privacy, regulation and big data in Europe and China. *Economy and Society*, 49(2), 187–212.
Andrejevic, M. (2007). *iSpy: Surveillance and Power in the Interactive Era*. Lawrence, KS: University Press of Kansas.
Baca, M. (2008). Introduction to Metadata Version 3. *Getty Information Institute*. http://www.getty.edu/publications/intrometadata/ [Accessed 21 August 2021].
Bakamo Social. (n.d.). Bakamo Public. https://www.bakamosocial.com/bakamopublic [Accessed 25 March 2022].
Bakamo Social. (2017). Summary of the French Election Social Media Landscape Report 2017. https://www.bakamosocial.com/wp-content/uploads/2021/03/BakamoSocial_FrenchMedia_StudySummary.pdf [Accessed 25 March 2022].
Bartlett, J., Smith, J., & Acton, R. (2018). *The Future of Political Campaigning*. London: Demos. https://demosuk.wpengine.com/wp-content/uploads/2018/07/The-Future-of-Political-Campaigning.pdf [Accessed 21 August 2021].
Bashyakarla, V., Hankey, S., Macintyre, A., Rennó, R., & Wright, G. (2019). *Personal Data: Political Persuasion. Inside the Influence Industry. How it works.* Tactical Tech, p.14. https://cdn.ttc.io/s/ourdataourselves.tacticaltech.org/Personal-Data-Political-Persuasion-How-it-works_print-friendly.pdf [Accessed 21 August 2021].
Bernal, P. (2016). Data gathering, surveillance and human rights: Recasting the debate. *Journal of Cyber Policy*, 1(2), 243–264.
Binney, W. (2016). Evidence to the Joint Parliamentary Committee on the Investigatory Powers Bill. JPCIPB. http://www.parliament.uk/documents/joint-committees/draft-investigatory-powers-bill/writtenevidence-draft-investigatory-powers-committee.pdf [Accessed 21 August 2021].
Bloomberg. (2020). Big data and business analytics market size is projected to reach USD 512.04 billion by 2026. *Valuates Reports*, 11 February 2020. https://www.bloomberg.com/press-releases/2020-02-11/big-data-and-business-analyt ics-market-size-is-projected-to-reach-usd-512-04-billion-by-2026-valuates-reports [Accessed 21 August 2021].
Bowen, P., Hash, J., & Wilson, M. (2006). *Information Security Handbook: A Guide for Managers*. Gaithersburg: National Institute of Standards and Technology (NIST).
Brandwatch. (n.d.). *Understand your Customers*. https://www.brandwatch.com/case-studies/ [Accessed 21 August 2021].
Brisco, K. (2021). Cyber threat basics, types of threats, intelligence & best practices. *Secure Works*. https://www.secureworks.com/blog/cyber-threat-basics [Accessed 25 March 2022].
Brodkin, J. (2015). Websites can keep ignoring "do not track" requests after FCC ruling. *Ars Technica*, 6 November 2015. https://arstechnica.com/information-technology/2015/11/fcc-wont-force-websites-to-honordo-not-track-requests [Accessed 21 August 2021].
Buchan, R. J. (2016). The international legal regulation of cyber espionage. *Tallinn Papers*, 65–86.
Burgess, M. (2017). The Emmanuel Macron email hack warns us fake news is an ever-evolving beast. *Wired UK*, 8 May 2017. https://www.wired.co.uk/article/france-election-macron-email-hack [Accessed 21 August 2021].
Center for Democracy and Technology. (2011). *Data Retention Mandates: A Threat to Privacy, Free Expression and Business Development* (October 2011). https://

cdt.org/insights/data-retention-mandates-a-threat-to-privacy-free-expression-and-business-development-1/ [Accessed 21 August 2021].
Christl, W., Kopp, K., & Riechert, P. U. (2017). Corporate surveillance in everyday life. *Cracked Labs*, 6, 2017-10.
Cinnamon, J. (2017). Social injustice in surveillance capitalism. *Surveillance & Society*, 15(5), 609–625.
Court of Justice of the European Union. (2014). Decision of 8 April 2014, Case C-293/12, Digital Rights Ireland v. "Ministry for Communications" et al. Luxembourg: Publications Office of the European Union.
Criteo. (n.d.) *Shopper Graph*. https://www.criteo.com/technology/shopper-graph/ [Accessed 21 August 2021].
Danaher, J. (2014). Rule by algorithm? Big data and the threat of algocracy, *Philosophical Disquisitions* (26 January 2014). http://philosophicaldisquisitions.blogspot.com/2014/01/rule-by-algorithm-big-data-and-threat.html [Accessed 21 August 2021].
Deibert, R. J. & Pauly, L. W. (2019). Mutual entanglement and complex sovereignty in cyberspace. In: Bigo, D., Isin, E., and Ruppert, E. (eds.), *Data Politics*. London: Routledge, pp.81–99.
Demarest, G. D. (1996). Espionage in international law. *Denver Journal of International Law and Policy*, 24 (1996), 321–326.
Determann, L. & Guttenberg, K. T. (2014). On war and peace in cyberspace security, privacy, jurisdiction. *Hastings Constitutional Law Quarterly*, 41(4), 875–902. https://repository.uchastings.edu/hastings_constitutional_law_quaterly/vol41/iss4/4/ [Accessed 21 August 2021].
*Deutscher Bundestag.* (2011). Drucksache 17/ 6233, Deutscher Bundestag, 17. Wahlperiode 8 (2011). http://dipbt.bundestag.de/ dip21/btd/17/062/1706223.pdf [Accessed 21 August 2021].
Diakopoulos, N. (2016). Computational journalism and the emergence of news platforms. In: Franklin, B. and Eldridge II, S. (eds.), *The Routledge Companion to Digital Journalism Studies.* London: Routledge, pp.176–184.
Digital, Culture, Media and Sport Committee. (2019). *Disinformation and "fake news": Final Report*. UK Parliament. https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf [Accessed 21 August 2021].
Driss, O. B., Mellouli, S., & Trabelsi, Z. (2019). From citizens to government policy-makers: Social media data analysis. *Government Information Quarterly*, 36(3), 560–570.
Enberg, J. (2019). What's shaping the digital ad market. *Insider Intelligence*. https://www.emarketer.com/content/global-digital-ad-spending-2019 [Accessed 21 August 2021].
EU Data Protection Supervisor (EDPR). (n.d.). Homepage. 2020. https://edps.europa.eu/data-protection/our-work/subjects/surveillance_en [Accessed 21 August 2021].
EU. (2016). *Social Media Platforms and the Digital Single Market Opportunities and Challenges for Europe*. COM/2016/0288 final. Luxembourg: Publications Office of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0288&from=EN [Accessed 25 March 2022].
European Commission. (2018). *The GDPR: New Opportunities, New Obligations*. Luxembourg: Publications Office of the European Union, p.3. https://ec.europa.eu/info/sites/info/files/data-protection-factsheet-sme-obligations_en.pdf [Accessed 21 August 2021].

Experian. (n.d.-a). *Consumer View: Data by the Numbers*. https://www.experian.com/assets/marketing-services/infographics/consumerview.pdf [Accessed 21 August 2021].
Experian. (n.d.-b). *Key Financial Data*. https://www.experianplc.com/investors/key-financial-data/ [Accessed 21 August 2021].
Facebook. (n.d.). About custom audience. Facebook.com https://www.facebook.com/business/help/744354708981227?id=2469097953376494 [Accessed 25 March 2022].
Falahrastegar, M. (2014). The murky world of third party web tracking. *MIT Technology Review*. https://www.technologyreview.com/2014/09/12/171400/the-murky-world-of-third-party-web-tracking/ [Accessed 21 August 2021].
Federal Trade Commission. (n.d.). Online tracking. https://www.consumer.ftc.gov/articles/0042-online-tracking [Accessed 21 August 2021].
Felten, E.W. (2013). Declaration in ACLU v Clapper and others, Case No. 13-cv-03994 (WHP) at the United States District Court, Southern District of New York.
Follorou, J. & Johannès, F. (2013). Révélations sur le Big Brother français. *Le Monde*. https://www.lemonde.fr/societe/article/2013/07/04/revelations-sur-le-big-brother-francais_3441973_3224.html [Accessed 21 August 2021].
Foulonneau, M. & Riley, J. (2014). *Metadata for Digital Resources: Implementation, Systems Design and Interoperability*. Amsterdam: Elsevier.
Fuchs, C. (2012). The political economy of privacy on Facebook. *Television & New Media*, 13(2), 139–159.
Gartner Glossary. (n.d.). Data Broker definition. https://www.gartner.com/en/information-technology/glossary/data-broker [Accessed 21 August 2021].
Gibbs, S. (2016). How much are you worth to Facebook. *The Guardian*. https://www.theguardian.com/technology/2016/jan/28/how-much-are-you-worth-to-facebook [Accessed 21 August 2021].
Google. (n.d.-a). About the customer matching process. *Google Ads Help*. https://support.google.com/google-ads/answer/7474263hl=en&ref_ topic=6296507 [Accessed 21 August 2021].
Google. (n.d.-b). Customer data: Definition. *Google Ads Help*. https://support.google.com/google-ads/answer/9004362?hl=en [Accessed 21 August 2021].
Greenberg, A. (2017). Everything we know about Russia's election-hacking playbook. *Wired*, 9 June 2017. https://www.wired.com/story/russia-election-hacking-playbook/ [Accessed 21 August 2021].
Greenwald, G. & MacAskill, E. (2013). NSA Prism program taps in to user data of Apple, Google and others. *The Guardian*.://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data [Accessed 21 August 2021].
Haggerty, K. D. & Ericson, R. V. (2000). The surveillant assemblage. *British Journal of Sociology*, 51(4), 605–622.
Harari, Y. N. (2018). *21 Lessons for the 21st Century*. London: Jonathan Cape.
IAB. (n.d.-a). Glossary: Digital media planning & buying. https://www.iab.com/wp-content/uploads/2016/04/Glossary-Formatted.pdf [Accessed 21 August 2021].
IAB. (n.d.-b). Understanding mobile cookies. https://www.iab.com/wp-content/uploads/2015/08/IABDigitalSimplifiedMobileCookies.pdf [Accessed 21 August 2021].
IAB and Winterberry Group. (2019). *The Outlook for Data 2019: A Snapshot Into the Evolving Role of Audience Insight*. https://www.iab.com/wp-content/uploads/2019/03/IAB_The-Outlook-for-Data-2019_2019-03-04_FINAL.pdf [Accessed 21 August 2021].

Investopedia. (2019a). The top ten technology companies. https://www.investopedia.com/articles/markets/030816/worlds-top-10-technology-companies-aapl-googl.asp [Accessed 21 August 2021].
Investopedia. (2019b). World top economies. https://www.investopedia.com/insights/worlds-top-economies/ [Accessed 21 August 2021].
Klamberg, M. (2010). FRA and the European Convention on Human Rights – a paradigm shift in Swedish electronic surveillance law. In: Schartaum, D. (ed.), *Nordic Yearbook of Law and Information Technology*. Bergen: Fagforlaget, pp.96–134.
Knapton, S. (2016). Facebook users have 155 friends – but would trust just four in a crisis. *The Telegraph*. https://www.telegraph.co.uk/news/science/science-news/12108412/Facebook-users-have-155-friends-but-would-trust-just-four-in-a-crisis.html [Accessed 21 August 2021].
Kruse, L. M., Norris, D. R., & Flinchum, J. R. (2018). Social media as a public sphere? Politics on social media. *Sociological Quarterly*, 59(3), 1–23.
Lazarus, D. (2019). Column: Shadowy data brokers make the most of their invisibility cloak. *Los Angeles Times* (5 November 2019). https://www.latimes.com/business/story/2019-11-05/column-data-brokers [Accessed 21 August 2021].
Lecher, C. & Brandom, R. (2016). Facebook caught an office intruder using the controversial surveillance tool it just blocked. *The Verge*. http://www.theverge.com/2016/10/19/13317890/facebook-geofeedia-social-media-tracking-toolmark-zuckerberg-office-intruder [accessed 21 August 2021].
Levine, B. (2016, 2 December). Report: What is data onboarding, and why is it important to marketers? *Martech Today*. https://martechtoday.com/report-data-onboarding-important-marketers-19292 4 [Accessed 21 August 2021].
Lewis, J. (2019). Economic impact of cybercrime. Center for Strategic & International Studies. https://www.csis.org/ analysis/economic-impact-cybercrime [Accessed 21 August 2021].
LiveRamp. (n.d.). Look-alike modeling: The what, why, and how. https://liveramp.com/blog/look-alike-modeling-the-what-why-and-how/ [Accessed 25 March 2022].
Lyon, D. (2003). Surveillance as social sorting: Computer codes and mobile bodies. In: Lyon, D. (ed.), *Surveillance as Social Sorting: Privacy, Risk, and Digital Discrimination*. London: Routledge, pp.13–30.
Lyon, D. (2007a). *Surveillance Studies: An Overview*. Cambridge: Polity Press.
Lyon, D. (2007b). Surveillance, security and social sorting: Emerging research priorities. *International Criminal Justice Review*, 17(3), 161–170.
MacAskill, E. & Dance, G. (2013). NSA files. *The Guardian*. https://www.theguardian.com/world/interactive/2013/nov/01/snowden-nsa-files-surveillance-revelations-decoded#section/1 [Accessed 21 August 2021].
Manheim, K. & Kaplan, L. (2019). Artificial intelligence: Risks to privacy and democracy. *Yale Journal of Law & Technology*, 21(1), 106–189.
Matsakis, L. (2018). The security risks of logging in with Facebook. *Wired*. https://www.wired.com/story/security-risks-of-logging-in-with-facebook/ [Accessed 21 August 2021].
Maycotte, H. O. (2015). Beacon technology: The where, what, who, how and why. *Forbes*. https://www.forbes.com/sites/homaycotte/2015/09/01/beacon-technology-the-what-who-how-why-and-where/ [Accessed 21 August 2021].
Mayer-Schönberger, V. & Ramge, T. (2018). *Reinventing Capitalism in the Age of Big Data*. London: Basic Books.
McDonough, J. (2018). Structural metadata & standardization failures: Just a little bit of history repeating. In *Balisage: The Markup Conference*. Symposium on Markup

Vocabulary Ecosystems, 30 July 2018. https://www.balisage.net/Proceedings/vol22/html/McDonough01/BalisageVol22-McDonough01.html [Accessed 25 March 2022].
Merkel, R. (2014). There's no such thing as privacy on the internet anymore. *The Washington Post*. https://www.washingtonpost.com/posteverything/wp/2014/07/28/theres-no-such-thing-as-privacy-on-the-internet-anymore/ [Accessed 21 August 2021].
Milano, M., O'Sullivan, B., & Gavanelli, M. (2014). Sustainable policy making: A strategic challenge for artificial intelligence. *AI Magazine*, 35(3), 22–35.
Miller, S. (2019). Consultants mostly optimistic on industry's future, but 2020 worries loom. *Campaigns & Elections*, 6 March 2019. https:// www.campaignsandelections.com/campaign-insider/consultants-mostly-optimistic-on-industry-s-future-but-2020-worries-loom [Accessed 21 August 2021].
Mitchell, M. (2019). Artificial intelligence hits the barrier of meaning. *Information*, 10(2), 51.
Mohit, A. (2019). Eliminating one of the biggest obstacles in business today: Mass data fragmentation. *Forbes*. https://www.forbes.com/sites/forbestechcouncil/2019/05/03/eliminating-one-of-the-biggest-obstacles-in-business-today-mass-data-fragmentation/ [Accessed 21 August 2021].
Montgomery, K. C. (2011). Safeguards for youth in the digital marketing ecosystem. In: Singer, D. G. and Singer, J. L. (eds.), *Handbook of Children and the Media* (2nd ed.). Thousand Oaks, CA: Sage Publications, pp.631–648.
Murray, D. & Fussey, P. (2019). Bulk surveillance in the digital age: Rethinking the human rights law approach to bulk monitoring of communications data. *Israel Law Review*, 52(1), 31–60.
Newberry, C. (2019). The Facebook pixel: What it is and how to use it. *Blog Hootsuite*. https://blog.hootsuite.com/facebook-pixel/ [Accessed 21 August 2021].
Novet, J. (2015). Facebook has paid out $8b to developers. *VentureBeat*, 25 March.
OECD. (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publications.
OssaLabs. (n.d.-a). Our Company. http://www.ossalabs.com/company [Accessed 25 March 2022].
OssaLabs. (n.d.-b). Social Media Monitoring for Political Campaigns and Organizations. http://www.ossalabs.com/political [Accessed 25 March 2022].
Revell, T. (2017). Hacking a US electronic voting booth takes less than 90 minutes. *New Scientist*. https://www.newscientist.com/article/2142428-hacking-a-us-electronic-voting-booth-takes-less-than-90-minutes/ [Accessed 21 August 2021].
Revell, T. (2018). How Facebook let a friend pass my data to Cambridge Analytica. *New Scientist*. https://www.newscientist.com/article/2166435-how-facebook-let-a-friend-pass-my-data-to-cambridge-analytica/ [Accessed 21 August 2021].
Richards, N. M. (2013). The dangers of surveillance. *Harvard Law Review*, 126(7), 1934–1965.
Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018). How Trump consultants exploited the Facebook data of millions. *The New York Times*. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html [Accessed 21 August 2021].
Rubinstein, I., Nojeim, G., and Lee, R. (2017). Systematic government access to private sector data: A comparative analysis. In: Cate, F. H. and Dempsey, J. (eds.), *Bulk Collection: Systematic Government Access to Private-Sector Data*. Oxford: Oxford University Press, pp.5–49.

Schuster, J. (2015). Political campaigns: The art and science of reaching voters. *LiveRamp*. https://liveramp.com/blog/political-campaigns-the-art-and-science-of-reaching-voters/ [Accessed 21 August 2021].
Schwartz, P. M. (2012). Systematic government access to private-sector data in Germany. *International Data Privacy Law*, 2(4, November), 289–301.
Schwartz, P. (2017). Systematic government access to private sector data in Germany. In: Cate, F. H. and Dempsey, J. (eds.), *Bulk Collection: Systematic Government Access to Private-Sector Data*. Oxford: Oxford University Press, pp.61–91.
Scott, J. C. (1998). *Seeing Like a State: How Certain Schemes to Improve the Human Condition have Failed*. New Haven, CT: Yale University Press.
Shields, R. (2014). The cross-device chasm and why statistical identification matters. *Exchange Wire*. https://www.exchangewire.com/blog/2014/04/22/the-cross-device-chasm-and-why-statistical-identification-matters/ [Accessed 21 August 2021].
Silverstein, J. (2019). Hundreds of millions of Facebook user records were exposed on Amazon cloud server. *CBS News*. https://www.cbsnews.com/news/millions-facebook-user-records-exposed-amazon-cloud-server/ [Accessed 21 August 2021].
Skinner, Q. (2013). Liberty, liberalism and surveillance: A Q&A with Quentin Skinner. *The Oxford University Politics blog*. https://blog.politics.ox.ac.uk/liberty-liberalism-and-surveillance-a-qa-with-quentin-skinner/ [Accessed 21 August 2021].
Smith, C. (2014). Reinventing social media: Deep learning, predictive marketing, and image recognition will change everything. *Business Insider*. http://www.businessinsider.com/social-medias-big-data-future-2014-3 [accessed 21 August 2021].
Spiegel. (2013). The German prism: Berlin wants to spy too. *Spiegel Online* (17 June 2013). https://www.spiegel.de/international/germany/berlin-profits-from-us-spying-program-and-is-planning-its-own-a-906129.html [Accessed 21 August 2021].
*The Economist*. (2017). The world's most valuable resource is no longer oil, but data. *The Economist* (6 May 2017). https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data [Accessed 21 August 2021].
Thurman, N. (2018). Social media, surveillance, and news work: On the apps promising journalists a "crystal ball." *Digital Journalism*, 6(1), 76–97.
Toor, A. (2015). France's sweeping surveillance law goes into e-ect. *The Verge* (24 July 2015). http://www.theverge.com/2015/7/24/9030851/france-surveillance-lawcharlie-hebdo-constitutional-court [Accessed 21 August 2021].
Treffer, P. (2019). Tory and National Front websites hid Facebook tracking pixel. *EUobserver*. https://euobserver.com/justice/141589 [Accessed 21 August 2021].
Trottier, D. (2011). A research agenda for social media surveillance. *Fast Capitalism*, 8(1). https://fastcapitalism.uta.edu/8_1/trottier8_1.html [Accessed 21 August 2021].
Trottier, D. (2016). *Social Media as Surveillance: Rethinking Visibility in a Converging World*. London: Routledge.
Trottier, D. (2020). Denunciation and doxing: Towards a conceptual model of digital vigilantism. *Global Crime*, 21(3–4), 196–212.
Tsukayama, H., Timberg, C., & Fung, B. (2016). Yahoo data breach casts "cloud" over Verizon deal. *Washington Post*. https://www.washingtonpost.com/news/the-switch/wp/2016/09/22/report-yahoo-to-confirm-data-breach-affecting-hundreds-of-millions-of-accounts/ [Accessed 21 August 2021].
Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*. https://firstmonday.org/article/view/4901/4097 [Accessed 25 March 2022].

Twitter. (2011). One million registered Twitter apps. Twitter blog, 11 July. https://blog.twitter.com/2011/one-million-registered-twitter-apps [Accessed 21 August 2021].
Watt, E. (2017). The right to privacy and the future of mass surveillance. *The International Journal of Human Rights*, 21(7), 773–799.
Wieczner, J. (2015). How investors are using social media to make money. *Fortune*, 7 December. http://fortune.com/2015/12/07/dataminr-hedge-funds-twitter-data/ [Accessed 21 August 2021].
Winston, M. (2017). Systematic government access to private sector data in France. In: Cate, F. H. and Dempsey, J. (eds.), *Bulk Collection: Systematic Government Access to Private-Sector Data*. Oxford: Oxford University Press, pp.49–61.
Wright, N. (2018). How artificial intelligence will reshape the global order. *Foreign Affairs* (10 July 2018). https://www.foreignaffairs.com/articles/world/2018-07-10/how-artificial-intelligence-will-reshape-global-order [Accessed 21 August 2021].
Wu, T. (2010). *The Master Switch: The Rise and Fall of Information Empires*. New York: Vintage.
YouGov Signal. (n.d.). Product. Take a tour. https://signal.yougov.com/product/ [Accessed 25 March 2022].
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.
Zuboff, S. (2016). Google as fortune teller, the secrets of surveillance capitalism. *Public Purpose* (5 March 2016). https://publicpurpose.com.au/wp-content/uploads/2016/04/Surveillance-capitalism-Shuboff-March-2016.pdf [Accessed 21 August 2021].
Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. London: Profile Books.

# 5. AI and the persuasion industry: Eroding the policy entrepreneurial resources and skills of citizens

## INTRODUCTION

> Computational politics turns political communication into an increasingly personalized, private transaction and thus fundamentally reshapes the public sphere, first and foremost by making it less and less public as these approaches can be used to both profile and interact individually with voters outside the public sphere (such a Facebook ad aimed at that particular voter, seen only by her). Overall, the impact is not so much like increasing the power of a magnifying glass as it is like repurposing the glass by putting two or more together to make fundamentally new tools, like the microscope or the telescope, turning unseen objects into objects of scientific inquiry and manipulation.
> (Tufecki, 2014)

Already in 1947, the founding father of public relations, Edward L. Bernays, argued that public consent toward state actions and decisions could be engineered, meaning that how citizens make decisions can be understood and precisely influenced. Based on the results of the propaganda efforts of states during the Second World War, in particular the UK government to keep the motivation of their citizens high, and the US government to change the minds of their citizens to accept US participation in the war efforts, Bernays realized that propaganda, which was then renamed public relations, can effectively influence citizens, within the limitations of democratic processes and values (Wu, 2016).

The efforts to influence citizens in Western democracies during the Second World War also showed that propaganda was effective when the same narrative was broadcasted on different channels, through different public figures, and in different contexts at the same time. Bernays argued that

> [t]he techniques can be subverted; demagogues can utilize the techniques for antidemocratic purposes with as much success as can those who employ them for socially desirable ends. The responsible leader, to accomplish social objectives, (…) must apply his energies to mastering the operational know-how of consent

> engineering, and to out-maneuvering his opponents in the public interest. (Bernays, 1947, p.115)

Until the last century, politicians won elections thanks to the same "traditional" communication tools as in previous centuries: meetings, rallies, media relations, printed press, etc. In the digital age, politicians must learn how to tweet regularly and develop their presence on social media platforms. Politicians and prominent citizens have always taken advantage of the power of information, and used it as an instrument to increase their support and suppress any form of dissidence. In the last century, broadcast communication enabled new forms of large-scale propaganda, including during wartime and in fascist regimes (Carson & Titcomb, 2017).

It is widely recognized that information and communication technologies (ICTs) have made an impact on how policy makers and citizens participate in the policy process (Chadwick, 2003). Computational politics encompasses all the activities such as outreach, persuasion, and citizen mobilization, which can now be conducted thanks to the analysis of large datasets (Tufecki, 2014). If using the personal data of clients to optimize the content and placement of ads is not new for businesses, it is more recent for political campaigns. Indeed, data-driven marketing is now commonly used by businesses but also for political campaigns. Data stemming from online user activity, behavioral data, demographics, psychological attributes, and predictive analytics allow political campaign leaders to identify patterns and predict future outcomes and trends, including what-if scenarios and risk assessment (Chester & Montgomery, 2017).

The broad use of digital technologies in politics led Farrell (2012) to argue that any political research must include an analysis of the role of the internet in this domain, since it is no longer possible to dissociate the internet and data from political practices today. Once primarily used for selling products and services to online users, data-driven political marketing and political campaigns are becoming the norm. These new practices have progressively been adopted in the context of elections, vote and referendums. Now, political leaders have indeed acquired new instruments to better understand citizens, tailor their messages based on this knowledge, and reach out to them individually and on a large scale.

Data-driven campaigning techniques were born in the US, where most digital technologies and the internet were developed. George W. Bush's second Presidential campaign (2004) and Barack Obama's two Presidential campaigns (Ambinder, 2009) already made use of big data collection and analytics to run their political campaigns and win the elections. More recently, the Cambridge Analytica scandal unveiled some concerning political persuasion practices performed on and by social media platforms in the context of

the Brexit and 2016 US Presidential election (Cadwalladr, 2018). It showed a "darker" side of social media platforms, whose machine learning algorithms (MLAs) are designed for profit maximization, not to bring people together, the well-being of users, nor the good of society (Alaimo & Kallinikos, 2017). Their surveillance capacity allows them to "control billions of minds every day" for profit (Harris, 2017); the human brain itself is now the subject of strategies to maximize profit (Carr, 2011).

The OECD (2001) argues that "[s]trengthening relations with citizens is a sound investment in better policy-making and a core element of good governance (…)" (p.1) since "it contributes to building public trust in government, raising the quality of democracy and strengthening civic capacity" (p.1). It also recommends that governments (1) enhance access to information so that citizens are well informed, (2) enable citizens to express their views on projects and societal issues that affect them in consultations, (3) engage citizens in decision-making processes. These three types of action (aka information, consultation, and active participation) must also be designed and implemented according to the principles of equity and inclusion, in order to avoid any discrimination within the population, and between the actors involved in the policy-making process.

As discussed in the second chapter, the relations between government and citizens span a wide range of interactions in the policy-making process (OECD, 2001). This chapter focuses mainly on the politics stream and the use of AI-powered computational tactics to influence citizens. It discusses how AI is used to profile users and potential voters online, and then examines AI-powered tactics used prior to and during political campaigns. This includes programmatic advertising, micro-targeting, A/B testing, smartphones and political apps, geotargeting citizens for political campaigns, false news and disinformation operations, social trolling and hybrid trolling, as well as automated profiles and social bots.

## PSYCHOMETRIC PROFILING

Psychometric profiling describes the ability to assess a person's psychological characteristics, including cognitive abilities, attitudes, personality, and knowledge abilities, according to a set of pre-defined criteria. The objective is to better understand how an individual thinks and behaves. The profiling is usually conducted through a self-assessment questionnaire or psychometric tests. The results allow the recruitment company for instance to ascertain if a potential candidate would be a good fit for the organization and the position he or she is applying for. Most of these psychometric tests are built on robust scientific research, which assessed large sample populations in order to define personality types. Such was the case with the Myers–Briggs Type

Indicator® (MBTI®)[1] that helps individuals identify their preferences among 16 personality types. It is used for individuals and organizations to evaluate job preferences, leadership capability, and emotional intelligence among other things. For instance, developed by Experian (2018), "Political Personas is a unique segmentation tool designed to help politicians, media owners, advertisers and agencies prepare for the seemingly never-ending campaign season by delivering a detailed understanding of key voter segments that go beyond party affiliation and political outlook."

Psychology has also been used extensively in past/recent decades by the advertising and public relations industries to target individuals with more precision and relevance, and thus influence their behavior. Thanks to the big data collected on the users of social media platforms, and the new forms of advertising they allow, psychometric profiling has gained new momentum. "Although there is nothing necessarily new about propaganda, the affordances of social networking technologies – algorithms, automation, and big data – change the scale, scope, and precision of how information is transmitted in the digital age" (Bradshaw & Howard, 2018, p.11).

Psychometric profiling and targeting for political purposes have become a concern among the general public and public officials since the emergence of the Cambridge Analytica scandal. Cambridge Analytica was the political consulting firm based in London, UK and New York, USA, which supported the Leave.UK party and the 2016 Trump campaign. In 2018, journalists from the British newspaper *The Guardian* unveiled a data breach scandal: the theft of personal data from over 50 million Facebook accounts in order to model user behavior online and target them during Brexit and the 2016 US Presidential election with bespoke messages, including false news (Cadwalladr, 2018). Later called the "Cambridge Analytica," from the name of the company that collected the data and offered its services to political parties and leaders around the world, this scandal shed light on a set of advanced data-driven practices to influence individual behavior that stem from commercial marketing innovations.

Cambridge Analytica was part of a "Bigger British company called SCL Group. It specialises in 'election management strategies' and 'messaging and information operations', refined over 25 years in places like Afghanistan and Pakistan. In military circles this is known as 'psyops' – psychological operations. (Mass propaganda that works by acting on people's emotions.)" (Cadwalladr, 2017).

Prior to the emergence of the scandal, the company boasted about its new psychographic profiling capacity, which was developed with the data collected on Facebook via an app, but without the consent of citizens. Data about over

80 million US citizens was stolen, and allowed the company to design their psychographic tool, based on the "big five" personality traits:

- openness to experience
- conscientiousness
- extraversion
- agreeableness
- neuroticism.

According to former CEO of Cambridge Analytica Alexander Nix (2013): "if you know the personality of the people you are targeting, you can nuance your messaging to resonate more effectively with those key audience groups. For a highly neurotic and conscientious audience, you're going to need a message that's rational and fear-based, or emotionally based."

In 2013, Kosinski, Stillwell and Graepel showed that Facebook likes could help predict sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender. For instance, their model correctly made the distinction between homosexual and heterosexual men in 88% of cases, African Americans and Caucasian Americans in 95% of cases, and between Democrat and Republican in 85% of cases. Several years later, Youyou, Kosinski, and Stillwell (2015) showed "that computers' judgments of people's personalities based on their digital footprints are more accurate and valid than judgments made by their close others or acquaintances (friends, family, spouse, colleagues, etc.)" and they consequently highlighted the fact "that people's personalities can be predicted automatically and without involving human social-cognitive skills" (p.1036).

In 2017, research from Matz, Kosinski, Nave, and Stillwell defined psychological mass persuasion as "the adaptation of persuasive appeals to the psychological characteristics of large groups of individuals with the goal of influencing their behavior." Their research reached over 3.5 million individuals with psychologically tailored advertising, and showed that "matching the content of persuasive appeals to individuals' psychological characteristics significantly altered their behavior as measured by clicks and purchases" (up to 40% more clicks and up to 50% more purchases than their mismatched or unpersonalized counterparts). Hence, their findings show "that the application of psychological targeting makes it possible to influence the behavior of large groups of people by tailoring persuasive appeals to the psychological needs of the target audiences" (p.12714).

The data collected and analyzed enable political marketeers to target citizens with individualized messages and across devices as discussed in the next sections.

## AI FOR ADS

Political campaigning benefits from the latest technological and marketing advances developed and used by the private sector to identify, target, and influence individuals to buy their products and services (Schuster, 2015). Social media platforms have provided political marketers and PR experts new tools and additional data to improve the effectiveness and efficiency of their marketing and persuasion techniques: digital technologies are increasingly used to mobilize voter turnout, raise funds, and support field engagement teams with data and content (Karpf, 2017). A new generation of political communication tactics and tools, encapsulated in the concept of computational politics, has emerged (Tufecki, 2014). Among them, new advertising tools powered by AI.

### Programmatic Advertising

In 2019, the year of the EU Parliamentary elections, European political parties launched digital marketing campaigns. The Statistica report contains aggregated statistics on spending and impressions about political ads on Facebook, Instagram, Google, and YouTube in the weeks leading up to the European Parliament election. For instance, the SPD (Social Democratic Party of Germany) spent almost 1.2 million euros on online political advertising in the three months leading up to the elections, followed by the Ciudadanos spending 885,000 euros, Podemos at 623,000 and Partido Popular (PP) at 586,000 (Statistica, 2019).

From 20 March 2019, ads that feature a political party, a current elected officeholder, a candidate for the EU Parliament, an elected national office within an EU Member State, or the UK Parliament are valued at 10,836,550 euros (Google, 2020) for a total of 138,367 political ads. But the number differs drastically from one country to another: Germany spent 944,650 euros while Spain spent 1,065,750 euros and Austria 1,032,700 euros for the same period but for about half the population. At the low end of the spectrum, France spent a total of 18,350 euros. To compare, the political ads on Google in the US reached 331,711,700 USD and a total of 390,107 ads from 30 May 2018 until 11 August 2020 (Google, 2020). In other words, although one more year has been included in the reporting for the US ads, the total spend is nevertheless 30 times greater (if we consider a US dollar–euro parity).

Twitter decided to ban all political advertising on their platform (Dorsey, 2019) to respond to the growing public pressure in times of political elections. At the EU level, as mentioned previously, social media platforms signed a voluntary code of conduct to render their political advertising policies and

practices more transparent, and tackle the dissemination of illegal content online (EU Commission, n.d.).

In terms of ad placement, new instruments allow 24/7 service without any human intervention. Traditionally, ad spaces were negotiated by humans and ads resulted from manual insertion orders. Today, programmatic advertising allows political campaign leaders to delegate these tasks to an AI that purchases digital advertising: "[w]hereas the traditional method includes requests for proposals, tenders, quotes and human negotiation, programmatic buying uses machines and algorithms to purchase display space" (Bilz, 2018). It works thanks to real-time auctions to buy and sell ads online. The growth of fake news is also associated with programmatic advertising since there is no longer any human control over selling and buying ads (Weissbrot, 2016).

> Programmatic's golden promise was allowing advertisers to efficiently buy targeted, quality, ad placements at the best price, and publishers to sell available space to the highest bidders…. What was supposed to be a tech-driven quality guarantee became, in some instances, a "race to the bottom" to make as much money as possible across a complex daisy chain of partners. With billions of impressions bought and sold every month, it is impossible to keep track of where ads appear, so "fake news" sites proliferated. Shady publishers can put up new sites every day, so even if an exchange or bidding platform identifies one site as suspect, another can spring up. (Clark, 2017)

Social media platforms and digital marketing companies offer programmatic advertising services, such as Facebook Ads Manager, AppNexus, which belongs to Xandr, the advertising and analytics division of the US telecom company AT&T, and Criteo. "Programmatic revenues in Europe grew by a healthy 33% in 2018 to reach €16.7bn, despite concerns from advertisers about the effectiveness of the current digital supply chain and worries that GDPR would impact digital spend" (Glenday, 2019).

### Micro-Targeting

Information and communication technologies (ICTs), once praised for providing broader access to information and empowering civil society, are now also recognized as global instruments of data collection, surveillance, and influence. This capacity to track the behavior of citizens is well encapsulated in the concept of surveillance capitalism (Zuboff, 2019):

> False consciousness is no longer produced by the hidden facts of class and their relation to production, but rather by the hidden facts of commoditized behavior modification. If power was once identified with the ownership of the means of production, it is now identified with ownership of the means of behavioral modification. (Zuboff, 2015, p.82)

In this age of surveillance capitalism, power is in the hands of the companies that have developed the capacity to collect and analyze big data. As described in the second chapter of this book, social media platforms have built large databases constituted of behavioral data, with the objective of transforming their users – citizens – into data subjects in order to manipulate their attitudes and decisions for profit (Aho & Duffield, 2020).

Opinion mining, sentiment analysis and digital listening describe these new practices that consist of tracking online conversations related to a topic or cause. Most of the time, they are performed on or by social media platforms. In addition to traditional polling and surveying, social listening allows companies and governments to "observe natural conversations without being part of it. This leads to a new perspective on reality – unfiltered truths straight from people" (Bakamo, n.d.). These tactics consist of first monitoring social media platforms to detect clusters of comments and opinions about a topic, cause or brand, in order to take appropriate measures (whether it is about proposing a new policy, adapting the political communication strategy, or shifting the positioning of a political leader) (Tran, 2020).

Data collected on social media platforms also enable emotion analytics: the content produced on social media platforms is analyzed by AI to "measure the impact of emotional advertising by assessing metrics like attention, emotional engagement and memory activation" (Jha & Ghoshal, 2019). Marketeers use AI to conduct a semantic and emotional content analysis on social media platforms to help them predict who will engage emotionally, and then target them accordingly (Kelshaw, 2017). This will allow digital marketers to target relevant individuals by "tailoring persuasive appeals to the psychological needs" (Matz, Kosinski, Nave, & Stillwell, 2017, p.12714).

AI predictions are based on past data and behaviors and AI algorithms learn from the large datasets collected. However, this can lead to biases (if the data collected contains biases) and/or can only reinforce previous patterns of behavior. AI can support more precise modeling of human behaviors, although with the restriction that the model assumes that individuals retain the same pattern of behavior over time, which is not the case. Human behavior can be irrational and change dramatically over time. Hence, AI cannot predict changes of opinion on topics such as climate change, that led to the emergence of new social movements such as FridaysforFuture[2] or Extinction Rebellion.[3] This limitation must be kept in mind, since it can lead to major errors in predictions.

Thanks to their knowledge about individual citizens, political marketeers can tailor the information inputs at the individual level with more and more accuracy. Campaign leaders can easily determine how and when to present an argument, and to whom: "[d]ata-driven campaigning gives you the edge that you need to convince swing votes one way or the other, and also to get certain people to show up to the polls" (Wakefield, 2019).

Micro-targeting indeed provides the ability to understand how an audience processes information at the individual level, and based on this knowledge, to then reach out to them with a highly individualized message. What is new here is the scope and the opacity of this process. As it takes place on social media platforms, personalized content can simultaneously reach millions of individuals, without the scrutiny and regulation that traditional media such as newspapers or TV channels must comply with.

The well-known case of Cambridge Analytica aptly illustrates this new, which places data collection and analysis at the heart of political campaigning: "[t]hrough smart use of machine learning, big data and extremely targeted and personalized digital ads, Vote Leave was able to identify key concerns facing the UK population and create campaigns that spoke directly to these concerns, targeting the right demographic of people for whom these concerns were most relevant" (Bender, 2017).

As mentioned, AI is now used in political campaigns to analyze large datasets, model the behavior of citizens according to psychographic and demographic indicators, and then make predictions for future behaviors. AI modeling precision is growing with the large datasets it is fed with. For the time being it is complementary to human strategists, but

> (…) might one day prove better than human strategists at working out exactly who should be targeted, when, and with what content, in order to maximize persuasive potential. AI would be capable of pulling together vast amounts of data from across different sources, and identifying relationships likely to remain invisible to human eyes. (Acton, 2018)

The following subsection will explore how smartphones and political apps are used for political marketing purposes. They are part of the computational politics tactics and tools since they also allow data to be collected and exploited in support of political campaigns.

## A/B TESTING

Another tactic to collect data about the preferences of users online is called A/B testing; it allows political marketers to collect data about the preferences of users toward specific content. This tactic is not new but was rapidly adopted by digital marketers since digital technologies make it easy to create two alternatives, make some changes and then see the impact on netizens in real time. It allows political marketers to assess the reaction of users (potential buyers or voters) to a message, a webpage, or any form of communication. It is commonly used by political parties to boost contributions to the political campaign (Bashyakarla, Hankey, Macintyre, Rennó & Wright, 2019). It is widely used

for websites, emails (subject lines, bodies), design elements (images, backgrounds, buttons), headlines, direct mail, TV, radio, phone, and even texting to "find the right messaging." The objective is to find the best advertising alternative that results in higher click-throughs.

Hatis is a pioneer in importing American campaign methods into Europe. The approach used is A/B testing. The "Vote Leave" Brexit party used A/B testing to improve the message targeted to a segment of the population:

> Cummings explained how the overarching theme of the Leave campaign was "Let's take back control". Based on research on public opinion of the EU, the campaign identified that "keep control" was an important theme. They then tested variations of this message – discovering that by including the word "back," in the messaging, they evoked rage in people. "Back" triggers loss aversion, playing into the strongly evolved instinct that we hate losing things, especially control. (Schneider, 2017)

In recent years, MLAs have allowed more automation of A/B testing services and the creation of new alternatives and combinations of text, image, and support.

However, A/B testing makes monitoring of political campaigns much more difficult: instead of checking one website, there are multiple variants of the same website to check. The same argument applies to micro-targeting on social media platforms (e.g. dark posts used during the Brexit campaign). Moreover, the increasing use of AI to develop new variants and messages may lead to the creation and online publication of political communication content with no human oversight.

## SMARTPHONES AND POLITICAL APPS

Political campaign apps are applications developed to support the campaign of a political candidate. Thanks to the use of smartphones and other GPS-enabled devices, political campaign leaders can follow citizens when they drive a car, shop at a store, or relax at home (Son, Kim, & Shmatikov, 2016). Contemporary political campaigns can follow targeted individuals cross-device (TV, websites, social media platforms, smartphones…), which augments their capacity to reach out at any time of the day.

Three types of political apps exist and contribute to political campaigns: mobile apps, enhanced canvassing apps, and games. Each of them is dedicated to a different audience as discussed below.

First, mobile apps are developed for supporters to obtain the latest information about their political candidate, and they provide an exclusive and dedicated space for them to interact. These apps also offer specific features, "such as letting users accrue points and unlock badges for completing certain tasks like watching campaign advertisements, tweeting pre-written political mes-

sages, sharing their contacts with the campaign or calling their representatives to discuss preset talking points" (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019).

The generalization of smartphones has changed how marketers target individuals. Indeed, citizens interact with information and make decisions differently because they have smartphones that provide easy and quick access to information: "What used to be our predictable, daily sessions online have been replaced by many fragmented interactions that now occur instantaneously. There are hundreds of these moments every day – checking the time, texting a spouse, chatting with friends on social media" (Ramaswamy, 2015a). Consequently, political advertising adapts its messages and communication format to this new way of consuming media, where citizens are repeatedly using a web-browsing device but for a specific intent: "Micro-moments occur when people reflexively turn to a device – increasingly a smartphone – to act on a need to learn something, do something, discover something, watch something, or buy something. They are intent-rich moments when decisions are made and preferences shaped" (Ramaswamy, 2015a). It is during these moments, when citizens search for information, look for entertainment while waiting for the bus or catching the next subway, that their opinions are progressively shaped: "There are no longer just a few sporadic 'a-ha!' moments of truth; now there are countless moments that matter" (Ramaswamy, 2015b).

Second, enhanced canvassing apps support the political campaign in two ways. First, they provide information about the citizens they visit, including their party registration, if they have voted in the past, the issues they care most about, and they also provide them tailor-made script and questions to ask. Second, canvassers upload information about the citizens they visit on the app, which provides the political campaign team with real-time and fresh information about citizens and potential voters. It is two-way information sharing.

Third, games or gamified apps contribute to promoting a political position in the run-up to an election (Bossetta, 2019). Examples include:

- SoBoHaZem Invaders to support Social Democratic party leader Bohuslav Sobotka in the Czech Republic (Švelch & Štětka, 2016);
- Pussywalk I and II to satirically penalize the public blunders of Czech President Miloš Zeman;
- Missione Bari in 2019 to support the mayoral reelection of Antonio Decaro in the Italian city of Bari;
- Super Klaver and Super Gruene developed to support the Green parties respectively in the Netherlands and in Germany (Bossetta, 2019). These are updated versions of the well-known Super Obama World, a remake of the 1991 Nintendo classic Super Mario World (Milian, 2008). In this new version, players control Dutch Groenlinks party leader Jesse Klaver (Super

Klaver) and the German Die Gruene party's leaders Katrin Goering-Eckardt or Cem Özdemir (Super Gruene) to shut down the coal-fired power plants and avoid political enemies;
- Fiscal Combat, a video game developed by the supporters of the far-left 2017 French Presidential candidate Jean-Luc Mélenchon, in which players roam the streets to shake money from the oligarchs and rival politicians (such as Emmanuel Macron and François Fillion, French politician Jérôme Cahuzac, who was prosecuted for tax evasion, former French President Nicolas Sarkozy, former head of the International Monetary Fund Christine Lagarde, and business leaders Pierre Gattaz and Liliane Bettencourt) to pay for Mélenchon's policies; once there are no more oligarchs, the player can see Mélenchon as he screams "Hypocrites!";
- Corbyn Run! to support candidate Jeremy Corbyn, is a video game similar to Fiscal Combat, where players knock down bankers, tax dodgers, Members of Parliament, and Conservative politicians (such as Theresa May, Boris Johnson, Jeremy Hunt, and even Margaret Thatcher) to invest in social services. Once an "enemy" is hit, the player receives money, which can then be released to support "manifesto pledges" that correspond to the core policy aspects of the Labour Party 2017 election manifesto (Dallison, 2017).

These apps tend to attract people who think alike, reinforcing the filter bubble and confirmation bias already present on social media platforms. In addition, because these apps do not stop functioning at the end of a campaign, they could contribute to creating a permanent campaigning context. The following section explores how geotargeting services are used for political marketing purposes.

## GEOTARGETING CITIZENS DURING POLITICAL CAMPAIGNS

Political marketers have crafted and delivered distinct messages to various segments of populations for a long time (e.g. urban vs. countryside, blue vs. white collar neighborhoods). With the rapid adoption of smartphones in Western democracies, it is possible to identify and target citizens on the go, and wherever they go throughout the day (Son, Kim, & Shmatikov, 2016), which triggered the emergence of a new set of digital marketing tools that benefit from this enhanced geolocalization-based targeting capacity (Warrington, 2015).

Geotargeting offers a broad range of techniques that enable marketers to deliver specific content to a user based on his geographic location, through GPS or Bluetooth signals, IP addresses, and more. Geotargeting is based on location data derived from "self-reported forms, publicly available voter rolls, private companies and data brokers, location-enabled services, APIs connected

to location-based apps, data licensed from third-party providers and more" (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019).

Geotargeting can provide in-depth information about what a person does, where and when during his day. It can provide data about trends and habits. In the context of a political campaign, this wealth of information about the "real" activities of citizens is a gold mine since it increases the precision of what marketers know about citizens, and consequently of their targeting strategies. From Liegey Muller Pons to Spallian, via Fédéravox and NationBuilder, electoral software and AI algorithms could well tip the next political campaigns.

Geotargeting was first used to push down messages to users based on their location: for instance, a person would pass a store and he would receive a text message with an ad. The well-known Augmented Reality (AR) game Pokémon Go that attracted millions of users at one point partnered with companies such as Starbucks to advertise their products to game users who were nearby (Schiff, 2017).

Three main forms of geotargeting techniques exist: mobile and property geotargeting, IP targeting, and geofencing.

First, mobile and property geotargeting allows marketers to target individuals with political messages either in print version sent by mail to their postal address, or by digital ads to their mobile phone (Schiff, 2017). Second, IP targeting allows marketers to target their ads and messages to specific individuals and organizations based on their Internet Protocol (IP) address. For instance, the political campaign team provides a list of names and addresses of individuals, and the IP targeting company matches these names with their IP addresses, and then displays the ads or video banner ads on their screens (Syzdek, 2017). Third, geofencing enables marketers to target users with a message when they enter a specific location: "geofence is a virtual polygon that can be placed around the perimeter of a location based on latitude and longitude coordinates" (Schiff, 2017).

As an illustration, the social media platform Snapchat introduced geofilters: "Geofilters are special overlays for Snaps that can only be accessed in certain locations – available based on your geolocation. Geofilters are a fun way to share where you are through filter overlays. They are specific to neighborhoods and special locations" (Snapchat, n.d.). In other words, geofilters offer the possibility to the user of sharing his location with others in an arty way. In 2015, Snapchat expanded geofiltering to ads through "sponsored geofilters," which provided brands with the possibility of creating their own arty filter that users could add to their snaps (or pictures) when visiting the brand store or restaurant (e.g. McDonald's cheeseburgers and pouches of fries, among others) (Chamberlain, 2016).

Geolocalization data is valuable for both targeting and content. First, data such as postal code or location trace at a gym is valuable to political marketers,

since it increases the precision of targeting. Citizens receive messages at a specific location and time. However, when combined with other data (for instance an organic supermarket they visit regularly, or health complement they buy), then the marketer can make assumptions about the values and beliefs of the citizen, and consequently target them with tailored messages:

> In the final part of the campaign, we knew turning young voters out was vital. This was the motivation behind our move to build an online tool showing people where to vote, and the decision to deploy Snapchat to get the message out to its 12 million young users in the UK. The results were staggering: 7.3 million individual people viewed our message and over 780,000 used the tool to find out where to vote. (Gwynne, 2017)

Liegey Muller Pons (LMP) collaborated with various political parties in Europe, including La République en Marche (LREM) in France, several socialist parties (PS in France, PSOE in Spain, PS in Belgium), and the Green Party in Bavaria, Germany. It also contributed to the political campaigns of Corrado Passera in Italy and Anne Hidalgo in France (Richaud, 2016) among others. It offered data science services to political parties and candidates. More precisely, LMP identifies where the physical door-to-door campaigning of political party activists should focus in order to convince undecided voters. LMP also offers tools to optimize targeted SMS and mailing campaigns (Richaud, 2017).

In the 2017 French Presidential election, Macron's campaign partnered with LMP and CloudFactory "to equip campaign teams with highly-enriched maps to effectively target voters on the ground. Such maps allow politicians to better understand local dynamics and to direct their campaigns toward areas where they have the greatest chance to win votes" (Wilson, 2017). Two forms of data retrieval coexisted and enriched the maps developed by LMP in previous political campaigns in France. One form, called open source, worked with open-source data, INSEE and the site data.gouv.fr. The other retrieved data from data brokers and online platforms. From these two sources of data, it is possible to produce a number of maps of electoral results that could help guide candidates in future elections. Mathilde Aubinaud, political communication expert, declares: "il s'agit en fait de déterminer les bureaux de vote et les quartiers stratégiques pour y accentuer les efforts de campagne – phoning, portes à portes, tractages"[4] (Branche, 2018).

Thanks to their use of geolocalization data, Macron's campaign teams on the ground could choose which polling stations to focus on based on household data (such as family size, socio-economic indicators, political affiliations, and other public data) and visualize voter demographics, voting behavior, and electoral results on an interactive map, which helped the political candidate understand the local population on a deeper level (Wilson, 2017).

In France, the use of data is not only political, but it can also be of interest to communities to test the effectiveness of public policy. Once elected, public decision-makers can use the data to help highlight their decisions. For example, Hatis' Carata software[5] enables communities to send targeted postal letters following public policy decisions that affect them. For instance: A speed bump installed? The inhabitants of the area are informed. A restored crèche? The parents concerned are informed by letter. The interest is also financial: communities reduce their expenses by sending letters only to inhabitants concerned, and the information is transmitted more efficiently. In other words, the data becomes a tool for optimizing the transmission of information while reducing its cost (Branche, 2018).

In 2016, Hatis developed a door-to-door tool, Knockin, for the primary of the main right-wing party in France (Les Républicains) which was designed to allow selected activists to directly access on their smartphone a list of people to meet door-to-door. "The map marked each contact's address with a red dot, along with the resident's name. Canvassers approached the app's contacts at their homes and addressed them by name, leading to a public outcry over its invasiveness" (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019).

As discussed in this section, disinformation campaigns make great use of false news to advance their geopolitical strategies and operationalize their tactics. Other instruments include hybrid trolling as discussed below.

## FALSE NEWS AND DISINFO OPS

Fake news has become a popular term, used by politicians, journalists, and citizens to information and sources they judge incorrect. In 2017, this term was chosen as the word of the year (Flood, 2017) and can be defined as "false, often sensational, information disseminated under the guise of news reporting" (Collins Dictionary, n.d.). However, fake news is more than false news, which can be simply described as incorrect data or information. Fake news is a term that encompasses more than the incorrectness of information or its source. It has become a political weapon to refute, without any rational argumentation, views from opposed parties and progressively refute the role of the press in liberal democracies. Fake news is "becoming a mechanism by which the powerful can clamp down upon, restrict, undermine, and circumvent the free press" (Wardle & Derakhshan, 2017, p.5). In other words, fake news was adopted by some "politicians around the world to describe news organizations whose coverage they find disagreeable" (Wardle & Derakhshan, 2017, p.5).

Fake news is a synonym for lying press. In the past, some politicians have already attacked the press, for example, Richard Nixon's vice president Spiro Agnew, who in his famous 1970 speech, talked about the press as "nattering nabobs of negativism" (Sullivan, 2016). Today, in contemporary Europe, the

Patriotic Europeans against the Islamization of the West (PEGIDA) movement and the Alternative für Deutschland (AfD) party in Germany often add the adjective "lying" when they mention the press during public rallies. In France, the right-wing party Rassemblement National (RN) regularly portrays the press in similar terms. These attacks on the quality and the role of the press also directly target mainstream media journalists on social media platforms. Hence, fake news is a lot more than false news. This book has adopted the term "false" news to describe the use of false content as examined below.

Recent research has revealed that Russian disinformation operators published over 45,000 messages on social media platforms in the last 48 hours of the Brexit campaign. False news is not new. It has been used throughout history for entertainment purposes and to gain power. For instance, Procopius, the great historian of the 6th century AD, was the official historian of the Emperor Justinian. He published portraits, although in secret he published what he called the anectdota (Procopius, 1896) which means secret history in Greek. They contained gossip about the sexual life of the emperor, and they were distributed alongside the official history. Procopius illustrates an early example of false news. But it is part of a whole tradition of false news (Darnton, 2017).

Already in the 18th century there was concern about false news coming from abroad, including from London (Darnton, 2014). The English press at that time was very violent. For instance, Henry Beat, who founded the *Morning Post* in 1772, defamed everyone, including Marie Antoinette, saying that she had had an affair with an Englishman and that the latter was receiving gifts from the Queen. Another example was the gutter journalist Charles Théveneau de Morande, director of the libelous *Gazetier Cuirassé* (Darnton, 2010), who lived in London in the 18th century and was only interested in French affairs. In the *Gazetier Cuirassé* (Darnton, 2010), there were footnotes making fun of the information. One of the footnotes says: "half of the information in this article is true" (Darnton, 2009). It was up to the reader to decide which part.

Charles Théveneau de Morande wrote a publication on the Madame de Barry, a mistress of the King. Only the title of the work is known to us: "les Mémoires secrets d'une femme publique ou recherches sur les aventures de Mme la comtesse du Barry depuis son berceau jusqu'au lit d'honneur, enrichis d'anecdotes et d'incidents relatifs à la cabale et aux belles actions du duc d'Aiguillon."[6] (Burrows, 2006). To silence him and prevent the manuscript from being published, the French police plotted to move the journalist to the Bastille. However, Théveneau de Morande discovered the plan, and spread rumors that the evil French, agents of despotism, were trying to kidnap a brave journalist. These rumors led to riots, which prevented the plan from being executed. As a result, Pierre-Augustin Caron de Beaumarchais was sent to offer Théveneau de Morande a great deal of money in exchange for his silence, to

which he agreed, burning the two manuscripts in front of Beaumarchais. The French police were ready to do anything to cover it up. French diplomats based in London expressed their impotence. In the end, the foreign minister, Charles Gravier, comte de Vergennes, spent more time dealing with bad publicity/commotion/upheaval/turbulence coming from England than with political issues, such as the Treaty of Paris which gave the Americans their freedom in 1783 (Darnton, 2014).

On the other side of the ocean, information was already being instrumentalized during the American Revolution as Robert Parkinson (2016b) shows. For instance, John Adams wrote in 1769 that he was busy with "a curious employment. Cooking up Paragraphs, Articles, Occurrences etc. – working the political Engine!" (Adams, Diary 1, p.3523 cited in Bradley, 2012) to diminish royal authority in Massachusetts. He and other leaders of the American Revolution intended to manipulate the public opinion with fabricated and imposter content. Further examples include the Governor of New Jersey William Livingston, who wrote letters containing fabricated content, such as the fact that the King sent foreign soldiers to kill Americans (Livingston, 1777) or the false issue of the real Boston newspaper that Benjamin Franklin fabricated when he was ambassador in Paris (Parkinson, 2016a) printed and sent to friends and newspapers with a letter assuring them of its veracity. He then wrote to his friend Richard Price:

> The ancient Roman and Greek Orators could only speak to the Number of Citizens capable of being assembled within the Reach of their Voice: Their Writings had little Effect because the Bulk of the People could not read. Now by the Press we can speak to Nations; and good Books & well written Pamphlets have great and general Influence. The Facility with which the same Truths may be repeatedly enforc'd by placing them daily in different Lights, in Newspapers which are every where read, gives a great Chance of establishing them. And we now find that it is not only right to strike while the Iron is hot, but that it is very practicable to heat it by continual Striking. (Franklin, 1782)

These examples illustrate well that every society is an information society according to the medium of the time, whether it is a song, a libel, a satirical periodical, or a tweet. Each political system has tolerated false news differently, and false news has adapted to new technologies and continued existing under various forms. In other words, false news will undoubtedly continue to exist as long as humans do. But the current mix of political systems, education level, and technology is unique in history. Therefore, it is crucial to examine how recent technologies influence the dissemination of false news in European liberal democracies.

Wardle and Derakshan (2017) describe the current information ecosystem as information disorder, which encompasses three forms of misuse of infor-

mation. First, disinformation refers to the intent to use false information: here it implies that incorrect information or data is produced and diffused with the purpose of becoming detrimental to a person, a group, an organization, or a country. Contrary to disinformation, misinformation is the production and diffusion of incorrect information or data but without the intention to harm, or said differently without the knowledge that the information is wrong. Mal-information is the third form of misuse of information identified by Wardle and Derakshan (2017), and describes "strategic dissemination of the true facts with a negative intent" (Keller, Schoch, Stier, & Yang, 2020, p.260).

In considering false information, Wardle identified seven types, on a scale from least serious to most serious disinformation efforts (Wardle, 2017):

- satire, or parody,
- false connection,
- misleading content,
- false context,
- imposter content,
- manipulated content,
- fabricated content.

First, satire, or parody, can trigger an emotional response and can have a strong impact on the audience. But it does not hide the fact that it is a parody: it is a reconstruction of reality to make people laugh or think. But the intent is clear, as well as the fact that the content does not reflect reality. So, the audience knows it needs to consider this type of content differently. However, satire or parody can also be the opportunity to contribute to reinforcing conspiracy theories that some citizens believe: "Russian state-backed media can increase public engagement with conspiracy theories without directly supporting them" (Bright et al., 2020). For instance *Russia Today* (RT) ridicules false claims against the Bill and Melinda Gates Foundation about their role in global health, and describe in depth the allegedly forced vaccination campaigns that would have sterilized millions of women in Africa, and numerous young people in India (Furlong, 2020).

Second, false connection describes content or news articles, videos, any type of multimedia, where the title has little or nothing to do with the content. In an age of virality, where media outlets earn revenue through advertising and web traffic, a title is crucial to attracting an audience and increasing the number of clicks. It is a tactic to increase the visibility of content. The issue is that citizens consult most content on social media platforms through their newsfeeds, a collection of titles and links to more content. There is a high probability when using newsfeeds that you will not click on the link or on all links to check the content of the video or the article, which means that you will

rely on the message in the title. Of course, the intent here can be to disinform. But it can actually be misinformation: a catchy title that focuses on one aspect of the content or that slightly exaggerates the content can simply be a marketing tactic to increase visibility. The intent might not be to cause prejudice to the audience; it can be simply to acquire/attract more viewers and increase revenue.

Misleading content is wrong information used in a correct context, whereas false context is correct information used in an incorrect context. In other words, misleading content provides incorrect information in the right context. This is often used with data, graphs, and percentages. Either the percentage is slightly increased or decreased to falsely represent an aspect of reality (e.g. women earn 20% less than men on average in the UK), or the context is wrong (e.g. women earn 20% less than men on average in the EU). In other words, the context or the content is changed, but one part or sometimes most of it remains correct, making it very difficult for citizens to identify this as false news.

As examples of false context, CGTN reported that the Covid-19 virus originated in Italy according to an Italian scientist, who then counter argued that he had been cited out of context (Tang, 2020). Articles from Xinhua written in French and Spanish showcased how the international community praised the successful poverty-reduction effort of China, but framed it in a different context: as if it were part of China's response to Covid-19 (Xinhua News Agency, 2020b, 2020c).

Fifth, imposter content is about mimicking an official source of information, e.g. government or media news outlets such as CNN and BBC, to diffuse false content. The imposter content gains credibility thanks to its false source, which lures the audience into believing the false information is correct (e.g. polling data) (Young African Leaders Initiative, n.d.). A similar form of imposter content is called astroturfing, where the real source of the content is hidden (e.g. PR agency, political party) to make the audience believe it comes from grassroots civil society organizations, thereby granting a different legitimacy to the message (Cho, Martens, Kim, & Rodrigue, 2011). It can be used to give a false image of public consensus about an issue or topic (Howard, 2003). Keller, Schoch, Stier, and Yang (2020) describe political astroturfing as the hijacking of political debate by state-sponsored digital platform users who pretend to be regular citizens. Astroturfing can contribute to changing public opinion and generating enough doubt to inhibit action (Lyon & Maxwell, 2004) as for example in the case of climate change. A case in point, CGTN and Xinhua in Spanish support the claim of China's fast progress to find a vaccine, allegedly supported by the British medical journal, *The Lancet*, which recognized the initial clinical trials as "promising" (Xinhua News Agency, 2020a). CRI in French quoted a study of the Pasteur Institute in Paris

that allegedly supported the view of an "unknown" origin of the virus and the "locally-circulating" strain of the virus diffusion (CRI, 2020).

Sixth, manipulated content encapsulates all efforts to manipulate written content, imagery, or videos to deceive the audience, for example deep fakes: a video of a public figure with a voice over of someone else, which consequently attributes to the person something he/she has never said.

And lastly, the most sophisticated type of false news is fabricated content, which is completely new and false content created to deceive, such as a website with false content, false reports, and multimedia. To illustrate, CGTN published a video where Italians play the Chinese anthem and sing "Grazie China" (CGTN, 2020), which various independent researchers have declared to be manufactured.

Every day, citizens face this large array of false news, from parody to fabricated content. This topology reveals the fact that false news is a complex issue, not only when debating the impact of false news on individuals, but also when considering the variety of forms of false news. The next sections will provide some examples of these seven types of false information. To identify the different categories of false news, Wardle (2017) identified three criteria:

- the type of content produced,
- the intent behind the production of this content,
- the diffusion of this content: how it is diffused and spread throughout civil society.

## SOCIAL TROLLING AND HYBRID TROLLING

Disinformation campaigns make extensive use of trolls to provoke individual users, as well as to diffuse false news, conspiracy theories: "[o]nline trolling is the practice of behaving in a deceptive, destructive, or disruptive manner in a social setting on the internet with no apparent instrumental purpose" (Buckels, Trapnell, & Paulhus, 2014, p.97). Their intention is merely "to shock, enrage, scare, or threaten – or, simply, to emotionally provoke readers" (Buckels, Trapnell, & Paulhus, 2014, p.97). In other words, trolls intend to capture the attention of the netizens and social media users for as long as possible with the intention of harming them emotionally (Bishop, 2014). The content they use is merely a tool to achieve their objective to provoke (Hardaker, 2010). "Today, the trolls are disrupting public discourse by adopting extremist positions on both sides of the political spectrum thereby attempting to create divisions within Lithuanian society, often by exploiting already existing dividing issues" (Willemo, 2019).

Simultaneously, trolling serves a "grander" purpose (Van Reenen, 2014) as identified by journalist Shawn Walker who investigated Russia's "troll army"

"where hundreds of paid bloggers work round the clock in order to flood Russian internet forums, social networks and the comment sections of western publications with remarks praising the president, Vladimir Putin, and raging at the depravity and injustice of the west" (Walker, 2015). This second type of trolls, defined as hybrid trolls, "communicate a particular ideology and, most importantly, operate under the direction and orders of a particular state or state institution" (Spruds et al., 2016, p.10).

The opaque nature of hybrid trolling, combined with the free circulation of information as one of the core values of Western democracies, make any resistance against trolling nearly impossible (Spruds et al., 2016). "In that perspective, information operations using current communication systems, social networks or deliberately created propaganda portals conducted to undermine a state's sovereignty by spreading hatred, fear, resentment and bad blood are an immense power that is indefensible under current international legal and security regimens" (Schmidt, 2014, p.79).

Research shows that the current generation of trolls, contrary to those used in the 2016 US election, focuses more on sowing chaos to disrupt the public discourse and exploiting existing divisive issues (Willemo, 2019). The operation usually starts in fan groups on Facebook, where they contribute with content related to the topic of the group, such as movies or popular singers. Then they progressively include false news in the content they produce, by for instance alternating one post about a popular singer and one post containing false news. This allows the trolls to expose false news to a large number of people at once (Willemo, 2019).

Linvill and Warren (2018) examined the tweets published by the twitter user names released by the United States House Intelligence Committee in June 2018 as being associated with the Russian "troll farm" Internet Research Agency (IRA). The user names they examined were human-operated troll accounts. The analysis helped them identify five categories of hybrid trolls used for disinformation campaigns:

- Right trolls diffuse right-leaning populist messages and divisive content about mainstream and moderate Republican politicians;
- Left trolls spread socially liberal messages, divisive content about mainstream Democrat politicians, and discuss gender, sexual, religious, but mostly racial identity;
- Newsfeed-generated local news present themselves as US local news aggregators;
- Hashtag gamer-shared content is dedicated almost entirely to playing hashtag games: a popular game word on Twitter where one asks a question through a hashtag and others answer the implied question in a tweet (Haskell, 2015). In this case, some tweets were political;

- Fearmongers diffuse content about a fabricated crisis.

An employee of the IRA described the organized and systematic use of digital technologies for disinformation purposes as "some kind of factory that turned lying, telling untruths, into an industrial assembly line" (Troianovski, Helderman, Nakashima, & Timberg, 2018). This use of trolls can be characterized as "industrialized political warfare" (Linvill & Warren, 2018).

False news and trolls are part of the array of instruments used to disinform citizens and hamper the social imaginary of European liberal democracies. The next subsection explores another set of instruments on social media platforms: social bots.

## AUTOMATED PROFILES AND SOCIAL BOTS

Bots and trolls are omnipresent on social media platforms. The difference is that troll accounts have human operators, whereas bot accounts are computer operated (Linvill & Warren, 2018).

Both are used by a large array of actors to abuse human biases and vulnerabilities to provide a false image of reality, to alter the perception of netizens. In the context of disinformation, bots and trolls are used to intensify certain political trends and views, promote certain interests, acquire influence on social media, and diffuse false news (Aiello, Deplano, Schifanella, & Ruffo, 2014). By projecting a false representation of reality, for example increasing or decreasing the support for some political movements, they can be very useful to silence dissidents (Pamment, Nothhaft, & Fjällhed, 2018). But the use of bots and trolls for disinformation also has more secondary consequences: by sowing chaos and polarization, they make it challenging to distinguish between truth and falsehood, which leads people to distrust all information.

In recent years, social media platforms have developed advanced bots to better identify and neutralize disinformation bots and automated fake accounts, meaning that disinformation operators now either continue their activities on less resourceful platforms, better hide their actions, or generate more genuine-seeming interactions to avoid detection (Bradshaw & Howard, 2018). Consequently, malign actors are constantly fighting a battle against platform moderators to discover new ways to get around the security measures of large social media platforms (Willemo, 2019).

Social bots can be defined as "algorithmically controlled accounts that emulate the activity of human users but operate at much higher pace (…) while successfully keeping their artificial identity undisclosed" (Bessi & Ferrara, 2016). Some of their most common uses are to interact with other users of social media platforms, diffuse content with a specific hashtag, or produce new content (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). Already

during the 2010 US election, social bots were used by some candidates to support their candidacy: they generated thousands of tweets to direct web traffic to websites with false news (Ratkiewicz et al., 2011). Although this type of strategy is well known and often referred to as Twitter bombs, attribution remains a challenge (Kollanyi, Howard, & Woolley, 2016). Any organization or state with sufficient resources can use an army of social bots to support their narrative and political agenda on social media platforms.

Social bots are easy to use and deploy. Some tech blogs provide how-to instructions for basic social bots, whereas other websites provide additional technical resources for more sophisticated ones. In terms of capacity, social bots are most commonly employed to perform the following tasks automatically:

- Look for phrases/hashtags/keywords on Twitter and share them;
- Respond to tweets that meet a certain criterion;
- Follow users that tweet with a specific phrase/hashtag/keyword;
- Follow back users that have followed the social bot;
- Follow any users that follow a specified user;
- Add users tweeting to public lists;
- Look for content according to specific criteria on web search engines and post them, or link them to other users;
- Aggregate public sentiment on specific topics of discussion;
- Buffer and post tweets (Bessi & Ferrara, 2016).

To quantify the presence of social bots is a complicated task, but Bessi and Ferrara (2016) identified about 400,000 bots engaged in the political discussion about the 2016 US Presidential election; they generated about 3.8 million tweets, which represent about one fifth of the entire conversation. Bots are also used to give more visibility to a narrative. For instance, China pushed for a narrative that positively depicts its management of the Covid-19 pandemic as will be discussed further.

As examined in this section, disinformation campaigns should be considered as part of a grand strategy to challenge the common understanding of the benefits, relevance, and resilience of European pluralist democracies, and by doing so, contribute to a global geopolitical power play.

## CONCLUDING REMARKS

As discussed in this chapter, governments, political leaders and parties have now the possibility to use a new generation of AI-powered computational tactics and tools. The case of Cambridge Analytica illustrated how contemporary political campaigns use advanced digital technologies to influence the

vote of citizens in a European liberal democracy (Cadwalladr, 2018). But the influence of digital technologies on the political landscape extends far beyond the day of the vote. This is a world of perpetual communication and campaigning. The precision and opacity of AI-powered computational tactics and tools weaken the citizen–government relation. These new persuasion tactics and tools do not offer the transparency and accountability necessary to ensure trust in the information citizens consume online (i.e. information that is "complete, objective, reliable, relevant, easy to find and to understand" OECD, 2001, p.1) and in the providers of this information. This is particularly concerning since a pluralism of information sources is essential for the proper functioning of democracies (Sidjanski, 1979).

The fact that these costly tools are mainly in the hands of governments, political leaders and parties strengthen an asymmetry of power between civil society and governments. Power lies in the hands of those who hold data and benefit from the AI-powered computational tactics and tools. Once enough is known about how citizens think, what triggers their emotions, and what opinions they hold, based on the large amount of data online platforms and other data brokers collect, and if they also have the capacity to reach out to each individual with a personal message based on the knowledge collected, at a national scale, then governments, political leaders and parties are able to anticipate how to influence the way someone thinks and how to trigger a specific opinion, hence a specific vote. These tools are mainly developed and provided by big tech companies. Hence, governments, political leaders and parties invite big tech companies in their relation with citizens. The dependence of "the politics" to big tech raises many governance and ethics concerns. What is more, the AI tactics and tools have the same characteristics are described in Chapter 1, and consequently introduce a high degree of uncertainty and vulnerability on the citizen–government relation.

What will the future hold for data collection and political campaigns? As in the past, a good idea is to look at innovations in the for-profit sectors, since all tech innovations used for political campaigns were first developed and used to promote and sell products and services. Hence, some of the trends can be highlighted here.

The existing trend that will continue to develop is connectivity and data. With more and more wearables, such as connected watches, and more and more connected things (also called Internet of Things or IoT), we will generate more and more data, which will allow marketers and tech companies to create more data points to scrutinize individuals' habits, behaviors, and attitudes, and adapt their products and services to our needs and interests: this trend is described by industry leaders as "customer segments of one" (Acton, 2018) meaning that marketing campaigns will be based on micro-targeting at the individual level. Group segments are less relevant when one can target the

individual. This of course will also feed political campaigns and marketers' strategies to influence the political behavior of citizens.

Furthermore, recent developments in neuroscience, cognitive computing, data analytics, and behavioral tracking, are increasingly harnessed by digital marketers to more effectively trigger a pre-set reaction among an audience (Crupi, 2015). "The field of neuromarketing – sometimes known as consumer neuroscience – studies the brain to predict and potentially even manipulate consumer behavior and decision making" (Harrell, 2019). It measures neural activity and physiological proxies for brain activity such as eye movements, facial expressions that can provide information about the emotional responses to a stimuli. "Gaze and pupil dilation can reveal a decision before it's made. These two biomarkers may offer clues into the underlying biological processes at play in decision making" (Michael Platt cited by Berger, 2020). Arousal is measured through other physiological responses, such as heart rate, respiration rate, and skin conductivity. Measurement is done through fMRI (functional magnetic resonance imaging) and EEG (electroencephalogram). Recent scientific advances "have demonstrated that brain data can predict the future success of products more accurately than can traditional market research tools such as surveys and focus groups" (Harrell, 2019). These recent developments can provide a better understanding of how citizens process information and make a political decision.

Additionally, the use of AI algorithms will continue to develop, in particular to generate unique content for individual users, which could "lead to a stream of unique, personalized messages targeted at each voter constantly updated based on A/B testing" (Acton, 2018). However, the questions related to their impact on citizens remain open. Indeed, it is hard to predict how citizens will adapt in the future to personalized and dynamic ads. With the growing awareness of psychological profiling and targeting, and their increasing interference in the private sphere of citizens, a movement against ads and the personalization of these ads lead some users to ban ads from their devices. Moreover, a general rejection against the personalization of ads in politics is well recognized by research: "An awareness of logging and using data on media consumption to inform political messaging can lead to a chilling effect among voters" (Bashyakarla, Hankey, Macintyre, Rennó, & Wright, 2019).

## NOTES

1. For instance, the Myers–Briggs Type Indicator (MBTI) is one of the world's most popular personality tools. https://www.themyersbriggs.com/en-US/Products-and-Services/Myers-Briggs [Accessed 21 August 2021].
2. See the website of this social movement to combat climate change: Fridays for Future. https://fridaysforfuture.org/ [Accessed 10 August 2021].

3. See the website of this social movement to combat climate change: Extinction Rebellion. https://extinctionrebellion.uk/ [Accessed 10 August 2021].
4. It is in fact a matter of determining the polling stations and the strategic districts to accentuate the campaign efforts there – phoning, door to door, leafleting.
5. See the website of Carata software. https://www.carata.eu [Accessed 10 August 2021].
6. The secret memoirs of a public woman or researches on the adventures of Mme la Comtesse du Barry from her cradle to the bed of honor, enriched with anecdotes and incidents related to the cabal and the beautiful actions of the Duke of Aiguillon.

# REFERENCES

Acton, R. (2018). The hyper-personalised future of political campaigning. *CAPX*. https://capx.co/the-hyper-personalised-future-of-political-campaigning/ [Accessed 21 August 2021].

Aho, B. & Duffield, R. (2020). Beyond surveillance capitalism: Privacy, regulation and big data in Europe and China. *Economy and Society*, 49(2), 187–212.

Aiello, L. M., Deplano, M., Schifanella, R., & Ruffo, G. (2014). People are strange when you're a stranger: Impact and influence of bots on social networks. *arXiv preprint arXiv:1407.8134*.

Alaimo, C. & Kallinikos, J. (2017). Computing the everyday: Social media as data platforms. *The Information Society*, 33(4), 175–191.

Ambinder, M. (2009). Exclusive: How Democrats won the data war in 2008. *The Atlantic*. https://www.theatlantic.com/politics/archive/2009/10/exclu-sive-how-democrats-won-the-data-war-in-2008/27647/ [Accessed 10 August 2021].

Bakamo. (n.d.). How we do it. https://bakamosocial.com/how-we-do-it/ [Accessed 23 March 2022].

Bashyakarla, V., Hankey, S., Macintyre, A., Rennó, R., & Wright, G. (2019). *Personal Data: Political Persuasion Inside the Influence Industry. How It Works*. Berlin: Technology Collective. https://cdn.ttc.io/s/ourdataourselves.tacticaltech.org/Personal-Data-Political-Persuasion-How-it-works_print-friendly.pdf [Accessed 21 August 2021].

Bender, B. (2017). How vote leave used data science and A/B testing to achieve Brexit. *AB Tasty*. https://www.abtasty.com/blog/data-science-ab-testing-vote-brexit/ [Accessed 10 August 2021].

Berger, M. W. (2020). Gaze and pupil dilation can reveal a decision before it's made. *Penn Today*. https://penntoday.upenn.edu/news/gaze-and-pupil-dilation-can-reveal-decision-its-made [Accessed 10 August 2021].

Bernays, E. L. (1947). The engineering of consent. *The Annals of the American Academy of Political and Social Science*, 250(1), 113–120.

Bessi, A. & Ferrara, E. (2016). Social bots distort the 2016 U.S. presidential election online. *First Monday*. https://firstmonday.org/ojs/index.php/fm/article/view/7090/5653 [Accessed 10 August 2021].

Bilz, N. (2018). GDPR-complaint against the "online behavioral advertising" industry. Datenschutz Notizen. https://www.datenschutz-notizen.de/gdpr-complaint-against-the-online-behavioral-advertising-industry-2121320/ [Accessed 23 March 2022].

Bishop, J. (2014). Representations of "trolls" in mass media communication: A review of media-texts and moral panics relating to "internet trolling". *International Journal of Web Based Communities*, 10(1), 7–24.

Bossetta, M. (2019). Political campaigning games: Digital campaigning with computer games in European national elections. *International Journal of Communication*, 13, 3422–3443.

Bradshaw, S. & Howard, P. N. (2018). Online supplement to Working Paper 2018.1 Challenging truth and trust: A global inventory of organized social media manipulation. https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf [Accessed 23 March 2022].

Branche, P. (2018). Comment La Donnée Révolutionne-T-Elle La Politique? *Forbes*. https://www.forbes.fr/business/comment-la-donnee-revolutionne-t-elle-la-politique/ [Accessed 21 August 2021].

Bright, J., Au, H., Bailey, H., Elswah, M., Schliebs, M., Marchal, N., … & Howard, P. N. (2020). Coronavirus coverage by state-backed English-language news sources. Project on Computational Propaganda, Oxford, UK, Data Memo.

Buckels, E. E., Trapnell, P. D., & Paulhus, D. L. (2014). Trolls just want to have fun. *Personality and Individual Differences*, 67, 97–102.

Burrows, S. (2006). *Blackmail, Scandal and Revolution: London's French Libellistes, 1758–1792*. Manchester: University Press.

Cadwalladr, C. (2017). Robert Mercer: The big data billionaire waging war on mainstream media. *The Guardian*. https://www.theguardian.com/politics/2017/feb/26/robert-mercer-breitbart-war-on-media-steve-bannon-donald-trump-nigel-farage [Accessed 21 August 2021].

Cadwalladr, C. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The Cambridge Analytica files. *The Guardian*. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election [Accessed 21 August 2021].

Carr, N. G. (2011). *The Shallows: What the Internet is Doing to Our Brains*. New York: W. W. Norton & Company.

Carson, J. & Titcomb, J. (2017). What is fake news? Its origins and how it grew in 2016. *The Telegraph*, 12 January. http://www.telegraph.co.uk/technology/0/fake-news-origins-grew-2016/ [Accessed 23 March 2022].

CGTN. (2020). Italians play Chinese national anthem to thank China for its aid. *CGTN*, 15 March 2020. https://news.cgtn.com/news/2020-03-15/Chinese-national-anthem-played-during-Italian-balcony-concert--OSzcwkrIL6/index.html [Accessed 10 August 2021].

Chadwick, A. (2003). Bringing e-democracy back in: Why it matters for future research on e-governance. *Social Science Computer Review*, 21(4), 443–455.

Chamberlain, L. (2016). GeoMarketing 101: What are geotargeting? *GeoMarketing*. https://geomarketing.com/geomarketing-101-what-is-geo-targeting [Accessed 10 August 2021].

Chester, J. & Montgomery, K. C. (2017). The role of digital marketing in political campaigns. *Internet Policy Review*, 6(4), 1–20.

Cho, C. H., Martens, M. L., Kim, H., & Rodrigue, M. (2011). Astroturfing global warming: It isn't always greener on the other side of the fence. *Journal of Business Ethics*, 104(4), 571–587.

Clark, J. (2017). Fake news: New name, old problem. Can premium programmatic help? *Advertising Age*. http://adage.com/article/digitalnext/fake-news-problem-premium-programmatic/308774/ [Accessed 25 March 2022].

Collins Dictionary. (n.d.). Fake news. https://www.collinsdictionary.com/dictionary/english/fake-news [Accessed 10 August 2021].
CRI. (1 May 2020). Institut Pasteur: Le Covid-19 En France Ne Provient Pas Directement De Chine. http://french.cri.cn/news/world/686/20200501/461120.html [Accessed 10 August 2021].
Crupi, A. (2015). Nielsen buys neuromarketing research company Innerscope. *Advertising Age*. http://adage.com/article/media/nielsen-buys/298771/ [Accessed 10 August 2021].
Dallison, P. (2017). Jeremy Corbyn, the video game. *Politico*. https://www.politico.eu/article/jeremy-corbyn-run-video-game-general-election-uk-2017/ [Accessed 10 August 2021].
Darnton, R. (2009). *The Devil in the Holy Water, or the Art of Slander from Louis XIV to Napoleon*. Philadelphia, PA: University of Pennsylvania Press.
Darnton, R. (2010). *Le Diable dans un bénitier. L'art de la calomnie en France, 1650–1800*. Paris: Gallimard.
Darnton, R. (2014). *De la censure. Essai d'histoire comparée*. Paris: Gallimard.
Darnton, R. (2017). On retrouve tout au long de l'histoire l'équivalent de l'épidémie actuelle de "fake news". *Le Monde*. https://www.lemonde.fr/idees/article/2017/02/20/la-longue-histoire-des-fake-news_5082215_3232.html [Accessed 10 August 2021].
Dorsey, J. (2019). We've made the decision to stop all political advertising on Twitter globally. We believe political message reach should be earned, not bought. Why? A few reasons…. [Tweet]. https://twitter.com/jack/status/1189634360472829952 [Accessed 10 August 2021].
Experian. (2018). Consumer view data | Experian marketing services, 2 May 2018. https://www.experian.com/assets/dataselect/brochures/consumerview.pdf [Accessed 25 March 2022].
Farrell, H. (2012). The consequences of the internet for politics. *Annual Review of Political Science*, 15, 35–52.
Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104.
Flood, A. (2017). The word of the year. *The Guardian*. https://www.theguardian.com/books/2017/nov/02/fake-news-is-very-real-word-of-the-year-for-2017 [Accessed 10 August 2021].
Franklin, B. (1782). From Benjamin Franklin to [Richard Price], 13 June 1782. *Founders Archive*. https://founders.archives.gov/documents/Franklin/01-37-02-0299 [Accessed 10 August 2021].
Furlong, R. (2020). Russian State TV Repeats Bizarre, Baseless Claims About Bill Gates And COVID-19. RadiofreeEurope. Available at: https://www.rferl.org/a/russian-state-tv-repeats-bizarre-baseless-claims-about-bill-gates-and-covid-19/30585799.html [Accessed 28 April 2021].
Glenday. J. (2019). Programmatic revenues hit €16.7bn in Europe last year – up 33%. *The Drum*. https://www.thedrum.com/news/2019/09/11/programmatic-revenues-hit-167bn-europe-last-year-up-33 [Accessed 10 August 2021].
Google. (2020). Political advertising in the European Union and the United Kingdom. *Google Transparency Report*. https://transparencyreport.google.com/political-ads/region/EU [Accessed 10 August 2021].
Gwynne, A. (2017). Theresa May called a snap election, but we in Labour had Snapchat. No contest. *The Guardian*. https://www.theguardian.com/commentisfree/

2017/jun/15/theresa-may-snap-election-labour-snapchat-campaigning [Accessed 10 August 2021].
Hardaker, C. (2010). Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions. *Journal of Politeness Research*, 6(2), 215–242.
Harrell, E. (2019). Neuromarketing: What you need to know. *Harvard Business Review*. https://hbr.org/2019/01/neuromarketing-what-you-need-to-know [Accessed 10 August 2021].
Harris, T. (2017). How a handful of tech companies control billions of minds every day. *TED*. https://www.ted. com/talks/tristan_harris_the_manipulative_tricks_tech_companies_use_to_capture_your_attention [Accessed 10 August 2021].
Haskell, W. (2015). People explaining their "personal paradise" is the latest hashtag to explode on Twitter. *Business Insider*. http://www.businessinsider.com/hashtag-games-on-twitter-2015-6 [Accessed 10 August 2021].
Howard, Philip N. (2003). Digitizing the social contract: Producing American political culture in the age of new media. *The Communication Review*, 6(3), 213–245.
Jha, D. & Ghoshal, M. (2019). When emotions give a lift to advertising. *Nielsen Featured Insights*. https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/nielsen-featured-insights-when-emotions-give-a-lift-to-advertising.pdf [Accessed 10 August 2021].
Karpf, D. (2017). Will the real psychometric targeters please stand up? *Civicist*. https://civichall.org/civicist/will-the-real-psychometric-targeters-please-stand-up/ [Accessed 10 August 2021].
Keller, F. B., Schoch, D., Stier, S., & Yang, J. (2020). Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication*, 37(2), 256–280.
Kelshaw, T. (2017). Emotion analytics: A powerful tool to augment gut instinct. *Think with Google*. https://www.thinkwithgoogle.com/intl/en-gb/consumer-insights/emotion-analytics-powerful-tool-augment-gut-instinct/ [Accessed 10 August 2021].
Kollanyi, B., Howard, P. N., & Woolley, S. C. (2016). Bots and automation over Twitter during the first US presidential debate. *Comprop data memo*, 1, 1–4. http://politicalbots.org/wp-content/uploads/2016/10/Data-Memo-First-Presidential-Debate.pdf [Accessed 10 August 2021].
Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15), 5802–5805.
Linvill, D. L. & Warren, P. L. (2018). Troll factories: The internet research agency and state-sponsored agenda building. Resource Centre on Media Freedom in Europe. http://pwarren.people.clemson.edu/Linvill_Warren_TrollFactory.pdf [Accessed 10 August 2021].
Livingston, W. (1777). To George Washington from William Livingston, 15 February 1777. *Founders Archive*. https://founders.archives.gov/documents/Washington/03-08-02-0369 [Accessed 10 August 2021].
Lyon, T. P. & Maxwell, J. W. (2004). Astroturf: Interest group lobbying and corporate strategy. *Journal of Economics & Management Strategy*, 13(4), 561–597.
Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, 114(48), 12714–12719. http://www.michalkosinski.com/home/publications [Accessed 10 August 2021].

Milian, M. (2008). New "Super Obama World" game (think Super Mario with a new president). *LA Times blog*. https://latimesblogs.latimes.com/washington/2008/11/super-obama-wor.html [Accessed 10 August 2021].
Nix, A. (2013). *Cambridge Analytica – The Power of Big Data and Psychographics*. Concordia Conference. YouTube. https://www.youtube.com/watch?v=n8Dd5aVXLCc [Accessed 21 August 2021].
OECD. (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publications.
Pamment, J., Nothhaft, H., & Fjällhed, A. (2018). *Countering Information Influence Activities: A Handbook for Communicators*. Swedish Civil Contingencies Agency (MSB), 22. https://mycourses.aalto.fi/pluginfile.php/1203540/mod_resource/content/1/countering%20information%20influence%20activities%20handbook.pdf [Accessed 10 August 2021].
Parkinson, R. G. (2016a). Fake news? That's a very old story. *The Washington Post*. https://www.washingtonpost.com/opinions/fake-news-thats-a-very-old-story/2016/11/25/c8b1f3d4-b330-11e6-8616-52b15787add0_story.html [Accessed 10 August 2021].
Parkinson, R. G. (2016b). *The Common Cause: Creating Race and Nation in the American Revolution*. Chapel Hill, NC: UNC Press Books.
Procopius. (1896). *The Secret History of the Court of Justinian*. McLean, VA: IndyPublish.
Ramaswamy, S. (2015a). How micro-moments are changing the rules. *Think with Google*. https://www.thinkwithgoogle.com/marketing-resources/micro-moments/how-micromoments-are-changing-rules/ [Accessed 10 August 2021].
Ramaswamy, S. (2015b). Outside voices: Why mobile advertising may be all about micro-targeting moments. https://blogs.wsj.com/cmo/2015/04/08/outside-voices-why-mobile-advertising-may-be-all-about-micro-targeting-moments/ [Accessed 10 August 2021].
Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011, July). Detecting and tracking political abuse in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1). https://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850 [Accessed 10 August 2021].
Richaud, N. (2016). Cette start-up sur laquelle Emmanuel Macron s'appuie pour sa grande marche. *Les Echos*. https://business.lesechos.fr/entrepreneurs/idees-de-business/021895448979-cette-start-up-sur-laquelle-emmanuel-macron-s-appuie-pour-sa-grande-marche-210792.php [Accessed 10 August 2021].
Richaud, N. (2017). LMP lève 4 millions pour exporter son logiciel de campagne politique. *Les Echos*. https://business.lesechos.fr/entrepreneurs/financer-sa-creation/030965672308-lmp-veut-exporter-son-logiciel-de-campagne-politique-316583.php [Accessed 10 August 2021].
Schiff, A. (2017). 2017 marketer's guide to location data. *Ad Exchanger*. https://www.adexchanger.com/mobile/2017-marketers-guide-location-data/ [Accessed 10 August 2021].
Schmidt, N. (2014). Neither conventional war, nor a cyber war, but a long-lasting and silent hybrid war/Nikoli konvencni nebo kyberneticka, ale dlouhodoba a nenapadna hybridni valka. *Obrana a Strategie/Defence & Strategy*, 2014(2), 73–86. http://goo.gl/H3C2gH [Accessed 10 August 2021].

Schneider, B. (2017). How Vote Leave used data science and A/B testing to achieve Brexit. *AB Tasty blog AI & Automation*. https://www.abtasty.com/blog/data-science-ab-testing-vote-brexit/ [Accessed 25 March 2022].

Schuster, J. (2015). Political campaigns: The art and science of reaching voters. *LiveRamp*. https://liveramp.com/blog/political-campaigns-the-art-and-science-of-reaching-voters/ [Accessed 10 August 2021].

Sidjanski, D. (1979). *Europe Élections de la démocratie européenne*. Paris: Stanké.

Snapchat. (n.d.). Snapchat geofilters. http://snapchatguides.blogspot.com/2015/06/snapchat-geofilters.html [Accessed 10 August 2021].

Son, S., Kim, D., & Shmatikov, V. (2016). What mobile ads know about mobile users. *NDSS '16*. http://www.cs.cornell.edu/~shmat/shmat_ndss16.pdf [Accessed 10 August 2021].

Spruds, A., Rožukalne, A., Sedlenieks, K., Daugulis, M., Potjomkina, M., Tölgyesi, B., & Bruge, I. (2016). *Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia. Results of the Study*. Latvian Institute of International Affairs Riga Stradins University.

Statistica. (2019). Political groups in the European Union (EU): Online advertising spend 2019. https://www.statista.com/statistics/1034662/eu-political-groups-online-advertising-by-number-of-ads/ [Accessed 10 August 2021].

Sullivan, M. (2016). How Trump attacks the media, and why that distorts reality. *The Washington Post.* https://www.washingtonpost.com/lifestyle/style/how-trump-attacks-the-media-and-why-that-distorts-reality/2016/07/24/32a456ea-5014-11e6-a7d8-13d06b37f256_story.html [Accessed 25 March 2022].

Švelch, J. & Štětka, V. (2016). The coup that flopped: Facebook as a platform for emotional protest. *First Monday*, 21(1). https://journals.uic.edu/ojs/index.php/fm/article/view/6333 [Accessed 10 August 2021].

Syzdek, J. (2017). What is IP targeting and how to use it to build a successful brand? *Digital Marketing Group*. https://thinkdmg.com/what-is-ip-targeting-and-how-to-use-it-to-build-a-successful-brand/ [Accessed 10 August 2021].

Tang, D. (2020). Beijing twisted my words on coronavirus's Italian origin, says scientist Giuseppe Remuzzi. *The Times*, 2020. https://www.thetimes.co.uk/article/beijing-twisted-my-words-on-coronaviruss-italian-origin-says-scientist-giuseppe-remuzzi-6twwhkrvn [Accessed 10 August 2021].

Tran, T. (2020). What is social listening, why it matters, and 10 tools to make it easier. *Blog Hootsuite*. https://blog.hootsuite.com/social-listening-business/#whatis [Accessed 10 August 2021].

Troianovski, A., Helderman, R. S., Nakashima, E., & Timberg, C. (2018, 17 February). The 21st-century sleeper agent is a troll with an American accent. *The Washington Post*.

Tufecki, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*. https://firstmonday.org/article/view/4901/4097 [Accessed 25 March 2022].

Van Reenen, C. (2014). From trolling for newbs to trolling for Cheezburger: An historical analysis of the communicative strategies of trolling. In: Moser, D. and Dun, S. (eds.), *A Digital Janus: Looking Forward, Looking Back*. Leiden: Brill Publishers, pp.153–163.

Wakefield, J. (2019). Brittany Kaiser calls for Facebook political ad ban at web summit. https://www.bbc.com/news/technology-50234144 [Accessed 10 August 2021].

Walker, S. (2015). Salutin' Putin: Inside a Russian troll house. *The Guardian*, 2 April 2015. http://www.theguardian.com/world/2015/apr/02/putin-kremlin-inside-russian-troll-house [Accessed 10 August 2021].
Wardle, C. (2017). Fake news. It's complicated. *First Draft*. https://firstdraftnews.org/articles/fake-news-complicated/ [Accessed 25 March 2022].
Warrington, G. (2015). Tiles, proxies and exact places: Building location audience profiles. *LinkedIn*. https://www.linkedin.com/pulse/tiles-proxies-exact-places-building-location-audience-warrington [Accessed 10 August 2021].
Weissbrot, A. (2016, 20 June). MAGNA and Zenith: Digital growth fueled by programmatic, mobile and video. Ad Exchanger. https://adexchanger.com/agencies/magna-zenith-digital-growth-fueled-programmatic-mobile-video/ [Accessed 23 March 2022].
Willemo, J. (2019). *Trends and Developments in the Malicious Use of Social Media*. NATO Strategic Communications Centre of Excellence. https://www.stratcomcoe.org/trends-and-developments-malicious-use-social-media [Accessed 10 August 2021].
Wilson, C. (2017). French presidential campaign rolls to victory using geospatial AI. *Cloud Factory*. https://blog.cloudfactory.com/french-presidential-campaign-geospatial-ai [Accessed 10 August 2021].
Wu, T. (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. New York: Alfred A. Knopf.
Xinhua News Agency. (22 May 2020a). Urgente: Prueba de Vacuna de China Contra Covid-19 Muestra Resultados Prometedores: *The Lancet*. http://spanish.xinhuanet.com/2020-05/22/c_139080036.htm [Accessed 10 August 2021].
Xinhua News Agency. (22 May 2020b). China Destinará Más de 20.000 Millones de dólares a Combatir la Pobreza En 2020. http://spanish.xinhuanet.com/2020-05/22/c_139079820.htm [Accessed 10 August 2021].
Xinhua News Agency. (23 May 2020c). La Communauté Internationale Salue la détermination de la Chine à Atteindre Ses Objectifs de Réduction de la Pauvreté. http://french.xinhuanet.com/2020-05/23/c_139081905.htm [Accessed 10 August 2021].
Young African Leaders Initiative. (n.d.). https://yali.state.gov/misinformation-spotting-the-signs-of-imposter-content/ [Accessed 10 August 2021].
Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4), 1036–1040.
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.
Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs.

# 6. AI and the weaponization of information: Hybrid threats against trust between citizens and democratic institutions

## INTRODUCTION

> Although there is nothing necessarily new about propaganda, the affordances of social networking technologies – algorithms, automation, and big data – change the scale, scope, and precision of how information is transmitted in the digital age.
> (Bradshaw & Howard, 2019, p.11)

> The control of information and telecommunication infrastructure, with the ability to respond to cyberattacks and to ensure cybersecurity, offers real power and is already one of the most significant political, economic and technological issues of the twenty-first century.
> (Ghernaouti, 2013, p.ix)

Disinformation is not new. Roman emperors and the British and German commanders in the two World Wars harnessed this power against their adversaries (Bittman, 1981). Joseph Goebbels, the Nazis' propaganda expert, modeled their propaganda efforts around the "Big Lie," the idea that repetition makes information truer (Ramakrishna, 2018). Jean Oberle (2017) reported that the Germans did not use bombs on Paris but rather false news: the Germans were pushing the former "Petainists" to act. During the Cold War, the USSR developed the concept of Reflexive Control theory (Thomas, 2004) to describe their approach to disinformation. A milestone in modern disinformation history is the Cold War, when the tensions between the two blocs led states to set up professional organizations whose main role was to produce and broadcast deception (Bittman, 1981).

According to Facebook, disinformation is intentional, often strategic in the sense that it targets specific demographics, and embeds false stories and coordinated efforts from real and fake accounts to engage the audience (Bennett &

Livingston, 2018). Facebook adopted the following operational definition of disinformation:

> Disinformation – Inaccurate or manipulated information content that is spread intentionally. This can include false news, or it can involve more subtle methods such as false flag operations, feeding inaccurate quotes or stories to innocent intermediaries, or knowingly amplifying biased or misleading information. Disinformation is distinct from misinformation, which is the inadvertent or unintentional spread of inaccurate information without malicious intent. (Weedon, Nuland, & Stamos, 2017, p.5)

Traditionally, an electoral campaign fulfills its role of informing and raising the awareness of very broad segments of society (Sidjanski, 1979). Thornton (2015) refers to the notion of "information warfare" to describe a context where information is the weapon and the minds of citizens the new "battlefield" (Cavelty and Mauer, 2008). They consist of operations to polarize civil society, sow chaos in the population of another state, and thereby weaken the opponent. Many tools and tactics exist, including AI-powered social bots to influence online conversations. Although varying in resources and capabilities, many governments' armed forces and intelligence agencies "have developed aggressive external operations" (Deibert & Pauly, 2019, p.83).

The previous chapter introduced the notion of fake news and discussed disinformation in the context of national politics. This chapter focuses on disinformation campaigns from an international relations perspective, meaning how disinformation is weaponized by some states to gain power, destabilize and weaken other states (Golovchenko, Hartmann, & Adler-Nissen, 2018). These operations target the trust that citizens have placed in their institutions (governments, representative mechanisms, the press), and in the social dialogue (leading to polarization of society). Without trustworthy information, citizens (more so than civil society at large) are more vulnerable to manipulation and disengagement from democratic processes. The grand strategy behind the operations takes time to assess, and therefore to identify all the operations that are in fact either stemming from the same group of actors or have the same grand strategy.

> The factors that make this strategy so powerful are that this type of "warfare" is continuously ongoing and hard to detect. It is complicated to identify its source, particularly as more often than not it is waged from several sources simultaneously. And finally, such a warfare strategy penetrates all levels of society at a very low cost. Even if the audience does not necessarily believe in the planted information, the abundance of unvetted information of itself leads to a persistent distrust of public information and the media. (Spruds et al., 2016, p.8)

This strategy is operationalized through tactics and tools that include new forms of deception and image-manipulation activities (Molander, Riddile, Wilson, & Williamson, 1996) and a weaponization of social media platforms. AI and more precisely Machine Learning Algorithms (MLAs) of social media platforms play a key role in the diffusion of disinformation campaigns. These tactics are about

> (…) influencing the target audience's values and belief system, their perceptions, emotions, motives, reasoning, and ideally, their behaviour. It is (…) aimed at maintaining the support of the loyal; convincing the uncommitted and undermining the opposition. This is achieved through influencing people's perception of what is going on and, in turn, influencing their online and offline behaviour by playing on emotional and logical arguments drawn from conversations and history, and by tapping into an existing narrative. (Nissen, 2015, p.84)

This chapter examines the main characteristics of disinformation campaigns in the context of a geopolitical power play on cyberspace. It argues that the liberal democratic model is under attack by authoritarian regimes, which use disinformation campaigns to threaten the citizen–government relation. The role of AI in this power play is crucial: it is used both to spread disinformation (intentionally through automated bots or unintentionally through MLA of social media platforms) and to defend against disinformation (the main solution to face the information avalanche produced and consumed globally). This chapter also highlights the difficulty of governments of liberal democracies to ensure that citizens have access to information that is "complete, objective, reliable, relevant, easy to find and to understand" (OECD, 2001, p.1). Their dependence on big tech companies to police the online information environment is problematic since it grants private companies the role to censor content. It also highlights the challenge that government and public entities face when enforcing law and the protection of citizens.

This chapter first discusses the ongoing geopolitical power play on cyberspace with various disinformation strategies and tactics. It then examines disinformation operations to harm trust in democratic institutions and the news ecosystem. It also discusses specific disinformation campaigns that were conducted prior to and during the Covid-19 pandemic in Europe. Finally, it presents non-technological responses to disinformation campaigns.

## GEOPOLITICAL POWER PLAY ON CYBERSPACE

The well-known paper "Cyberwar is Coming!" by the two RAND Corporation scientists, John Arquilla and David Ronfeldt (1993), argued that the information revolution was altering not only how conflicts take place (e.g. parties involved, terrain, technologies, etc.), but also the nature of conflicts, spurring

a need for new military structures, doctrines and strategies. The use of disinformation as part of the tactics to weaken other states is not new: intelligence services have always led espionage and reconnaissance activities to support kinetic military operations. What is different today is the domain in which disinformation is spread. This domain is man-made, meaning that it can be easily altered and manipulated.

In the mid-2000s, the topic of cybersecurity was increasingly discussed by the media and policy makers, following cyberattacks such as the 2007 attack against Estonia. This cyberattack followed the relocation of a statue commemorating the engagement of the Russian army in the Second World War to the suburbs of the city. The choice of day to relocate this statue was not random, since it was 1 May, the same day Russia honors its participation in the Second World War. The cyberattack that followed targeted websites and online services of private and public organizations, including banking and government services, as well as newspapers and broadcast media. This attack didn't cause any human harm but millions of dollars of losses (McGuinness, 2017). Although Russia denied any involvement, Estonia pointed at its neighbor and international experts agreed (Van Puyvelde & Brantly, 2019).

This attack put cybersecurity on the agenda of a larger number of countries, since they realized that increased connectivity meant increased vulnerability as well, as pointed out by the US joint forces years later: "The prosperity and security of our nation are significantly enhanced by our use of cyberspace, yet these same developments have led to increased exposure of vulnerabilities and a critical dependence on cyberspace, for the US in general and the joint force in particular" (Joint Publication, 2018a, p.I–2). This also signaled the realization that cyberconflict can be ongoing in times of inter-state peace. States conduct offensive operations in cyberspace, whether it is to steal information, destroy capabilities or infrastructure, or disinform populations, when there is no conflict between them. Sheldon (2011) characterizes cyberspace both as a space where conflict can take place, as well as means to advance their interests: "Cyberspace is the domain in which cyber operations take place; cyberpower is the sum of strategic effects generated by cyber operations in and from cyberspace" (p.96).

Cyberspace is poorly regulated. A large-scale disinformation campaign to interfere with the internal affairs of another state, including electoral processes, does not constitute an act of war that could justify a kinetic military response. The diplomatic response given during the last days of the Obama administration to the Russian interference (CNN, n.d.) is a good illustration of this challenge: "After discovering the existence, if not the full scope, of Russia's election interference efforts in late-2016, the Obama Administration struggled to determine the appropriate response. Frozen by 'paralysis of analysis,' hamstrung by constraints both real and perceived, Obama officials

debated courses of action without truly taking one," said committee chairman Senator Richard Burr, a North Carolina Republican (Cohen & Herb, 2020).

Disinformation campaigns take place primarily on social media platforms designed and maintained by companies with headquarters in the USA, and most disinformation content comes from outside Europe. Consequently, this means that disinformation campaigns are influencing European citizens and states but on foreign media and with foreign content. This double territoriality challenge adds to the qualification challenge discussed previously. These elements challenge traditional military strategies and call for developing new military strategy specifically dedicated to disinformation. Moreover, a specific tactical warning system is required to distinguish between disinformation attacks and other activities such as espionage or accidents.

In his famous paper, Thomas Rid (2012) argued that "any act of war has to have the potential to be lethal; it has to be instrumental; and it has to be political" (p.5), which consequently means that cyberwar will never take place: "Cyber war has never happened in the past. Cyber war does not take place in the present. And it is highly unlikely that cyber war will occur in the future" (Rid, 2012, p.6). In 2013, NATO Cooperative Cyber Defence Centre of Excellence (CCDCOE) sponsored a research project that led to the publication of the Tallinn Manual, which addressed the most severe cyber operations, and where the authors argue that a cyberattack can only be considered an armed attack if its impact leads to injury, death, or destruction (Schmitt, 2013, p.106). Nevertheless, the existence of such conflict requires European pluralist democracies to not only adopt measures to combat disinformation campaigns, but also to perceive these cyber operations for what they are: cyberattacks against the integrity of a state, an interference in the internal affairs of a state, and an attempt to weaken European pluralist democracies. We will refer to the notion of cyberconflict that reflects the reality of the ongoing offensive and defensive operations, and yet avoids the challenging question of qualifying these operations.

The cyberspace layer model developed by the US military is helpful to distinguish between different types of targets, tactics and actors. This model identified three interrelated layers with specific actors, attacks, and technologies. The physical layer corresponds to the physical IT devices and infrastructure (e.g. computing devices, storage devices, network devices, and wired and wireless links). The logic network corresponds to the logical connections between network nodes. Finally the cyber-persona consists of users, whether human or automated, as well as the content created and their behaviour (Joint Publication, 2018b). Disinformation campaigns target the top layer of cyberspace (cyber-persona layer). Because governments, companies, and civil society all use these platforms, they can be directly or indirectly affected by these efforts. Since space and time fail to exist in cyberspace, "targets

in the continental United States are just as vulnerable as in-theater targets" (Molander, Riddile, Wilson, & Williamson, 1996, p.xiii).

However, as mentioned previously, these operations can also be part of a broader strategy that includes other forms of cyberattacks that target the two other layers. In this context, cyber operations should be understood in relation to other domains of warfare and human activity (Van Puyvelde & Brantly, 2019). It is particularly the case for hybrid conflicts, where online and offline offensive operations are simultaneously conducted:

> Disinformation campaigns, in particular by third countries, are often part of hybrid warfare, involving cyber-attacks and hacking of networks. Evidence shows that foreign state actors are increasingly deploying disinformation strategies to influence societal debates, create divisions and interfere in democratic decision-making. These strategies target not only Member States but also partner countries in the Eastern Neighbourhood as well as in the Southern Neighbourhood, Middle East and Africa. (EU, 2018g, p.3)

Cyberconflicts have specific characteristics that affect how and why disinformation campaigns are led in cyberspace. Lindsay (2013) recognizes three of them: First, critical economic and military infrastructure is highly vulnerable to cyberattacks, making developed states a prime target. Second, offense has become easier while defense is growing harder in cyberspace. This is due to anonymity and the fact that few organizations share information when they are attacked, that a cyber-weapon can be used, sold, and re-used numerous times before it is identified by the maker of the vulnerable technology, and that it can be patched, which makes cyberattacks fairly risk-free and accessible even without technology expertise. Third, traditional deterrence does not work in cyberspace: the attribution issue undermines deterrence. For these reasons, cyber-tools are often considered by cybersecurity scholars as "the weapon of the week" in the sense that they empower "weaker" states (i.e. with lower penetration rates and fewer military and economic resources), more so than "stronger" states (i.e. developed, military resourceful, and highly connected states) (Van Puyvelde & Brantly, 2019). In other words, connectivity should be considered simultaneously as an asset and a liability. It is precisely this dual characteristic of connectivity as an asset and a liability that makes European citizens vulnerable to disinformation campaigns.

## Disinformation Strategies

State-sponsored disinformation campaigns pursue grand strategies where information is used to reach political and military objectives (Thornton, 2015). As mentioned previously, strategy has to do with the direction of an organization (Johnson, 2017). Clausewitz argued that "[t]he strategist must

therefore define an aim for the entire operational side of the war that will be in accordance with its purpose" (Clausewitz, 2003, p.177). To illustrate a State-sponsored disinformation, this section focuses on Russia. However, it does not imply that it is the only state conducting disinformation operations abroad. As briefly mentioned in a previous chapter, the use of information and disinformation to manipulate populations is probably as old as civilization. The choice of Russia is linked to its long history of information weaponization, which leads to a greater accessibility of the sources and analysis.

Russian and Western scholars tend to use concepts such as "hybrid conflict," "new generation warfare," "the Gerasimov Doctrine," "cross-domain coercion," and "gray zone tactics" among others (Chivvis, 2017; Adamsky, 2015; Morris et al., 2019; Galeotti, 2018; Kofman, 2016). They aim to describe the Russian understanding that modern warfare must be conducted through armed violence as well as non-military tactics (Chekinov & Bogdanov, 2015a, p.34; Chekinov & Bogdanov, 2015b; Gerasimov, 2013; Burenok, 2018, pp.61–66).

The employment of non-military measures in warfare is not a new debate among Russian military elite. For instance, a 1920 Russian military manual stated that: "Political sentiment of the population in an enemy's rear plays a big role in an opponent's successful activities; because of this it's extremely important to generate sentiments among populations against the enemy and use them to organize people's uprisings and partisan detachments in the enemy's rear" (Shil'bakh & Sventsitskiy, 1927).

Yet, it is only since the early 2000s and the Ukraine crisis that senior Russian leaders and military theorists formed a consensus on this new conceptualization of warfare (Lilly & Cheravitch, 2020), where the line between war and peace cannot be so clearly distinguished anymore, and where the weaponization of information and cyberattacks can be as effective as violent measures (Jonsson, 2019). In 2011, Russia's Ministry of Defense provided a clear description of its intention to weaponize information in the context of conflicts:

> (…) inflicting damage to information systems, processes and resources, critical and other structures, undermining the political, economic and social systems, a massive psychological manipulation of the population to destabilize the state and society, as well as coercing the state to take decisions for the benefit of the opposing force. (Ministry of Foreign Affairs of the Russian Federation, 2011)

The 2014 Russian Military Doctrine (Russian Federation Security Council, 2014) identifies a series of geopolitical threats and the new methods that the West is deploying against the country. According to this doctrine, this new context is forcing Russia to adopt a new strategy that consists of military and non-military measures, and new and non-traditional methods (Darczewska,

2015) including information operations as a defense tool (Darczewska, 2015). Another policy document that refers to information security is the Russian Strategy for Counteracting Extremism: internet and online forums are identified as spheres of great importance for Russian security, since they can be utilized to foster ethnic, religious, and national hatred and organize terrorist activities. Most policy documents present Russia as a defensive actor (Spruds et al., 2016) and only refer to Russia's efforts to fight "for the demilitarisation of […] the global information network, because it cannot permit the country and its surrounding areas to come under American 'quasi-occupation'" (Darczewska, 2015). By positioning the country and its actions as a defense against the threats coming from the US, NATO and other allies, it allows Russian authorities to justify intervention in the information space of its own population. Also, the quasi-absence of any mention of offensive uses of information and psychological persuasion abroad comes from the fact that their value lies in their covert nature (Spruds et al., 2016), which renders the attribution even more challenging.

In their 2019 report "Warring Songs: Information Operations in the Digital Age," Krasodomski-Jones, Smith, Jones, Judson, and Miller identified four strategic aims of disinformation campaigns conducted abroad:

- influence sympathetic changes in citizen behavior and perception,
- reduce the participation of one part of the population in the decision-making process,
- decrease the quality of their communications environment,
- diminish the quality of information available to citizens.

These strategic aims correspond to two grand strategies: the first two target the trust citizens have in their institutions, whereas the last two focus on the trust citizens have in news gatekeepers. A single disinformation campaign often combines several of these strategies and tactics at once. These two grand strategies and the tactics used to support them are presented in the following two subsections and will be illustrated in the third section of this chapter by disinformation operations during the 2020 Covid-19 pandemic. The examples stem from two studies conducted by the Oxford Internet Institute for the Project on Computational Propaganda.[1]

## Disinformation Operators

Disinformation campaigns have specific characteristics. Molander, Riddile, Wilson, and Williamson (1996) identified seven key features. First, their low cost of entry allows a large number of actors to engage; an internet connection and a laptop are enough to sabotage or conduct malevolent activities. Second,

disinformation campaigns blur the lines between geographical spaces, public and private motives, warfare, and criminal conduct (Molander, Riddile, Wilson, & Williamson, 1996). They are blurry in terms of time, space, and grand strategy. Their duration is spread over a long period and is difficult to clearly define. The geographical space where they occur is not clearly delimited: they can take place cross-border by supporting opposing parties on both sides of a border for example, or by supporting the extreme political parties in several countries. And yet, although they blur the lines, they are connected to all the other domains of military operation (Gartzke, 2013), business, and society as mentioned previously.

The identification of the source of the actions is a challenge in cyberspace. This is the well-known issue of attribution and the anonymous nature of cyberattacks (Libicki, 2009). Without specific identification, hidden behind multiple fake identities, the high level of anonymity allows disinformation operators to spread messages in the dark. Yet, Bennett and Livingston (2018) identified four categories of actors who are the most susceptible of producing and diffusing false news:

- news outlets promoting radical right-wing agenda, anti-immigrant and globalist conspiracies;
- disinformation operations from foreign states targeting elections and governments;
- political parties and movements supplying party updates punctuated with "nostalgic" nationalist content, such as the Austrian Freedom Party;
- for-profit content producers that use false news to attract web traffic and make a profit from the attention economy.

Reporters Without Borders (2019) adds another category, which in fact consists of the first two identified by Bennett and Livingston (2018), and further defines "state-backed media outlets as organisations that are either directly funded by the state or are editorially managed by their governments." Reporters Without Borders considers China, Iran, Russia, and Turkey as the countries where governments are most likely to obstruct news reporting. This control of information is not limited to internal audiences, but also to international ones, where they use their national media outlets to diffuse disinformation. The state-backed media outlets considered in this chapter are the ones with a global outreach and which target their communication on Europe.

The largest English-speaking state-backed media outlets from China, Iran, Russia, and Turkey include:

- China: China Global Television Network (CGTN)*, China Daily, China Plus, the People's Daily, China Radio International (CRI)*, and Xinhua News Agency*

- Russia: RT* (formerly Russia Today) and Sputnik*
- Iran: Mehr News and Press TV*
- Turkey: Anadolu and TRT network.* (*Some of these outlets also have services in other European languages, such as German, French, and Spanish.)

To make the situation even more complex, the for-profit false news producers often take on partisan aspects (Bennett & Livingston, 2018), since these allow them to better target the message and substantially increase web traffic. This partisan for-profit content is then picked up by bots, which can be part of systematic disinformation strategies, by either national or foreign actors.

The apolitical disinformation entrepreneurs (Bennett & Livingston, 2018), create fake stories that circulate on social media platforms, and thanks to the algorithms, and the sensationalist content and good targeting, become viral and bring traffic to the political websites they own and manage, the objective being to make a profit from the advertising that is displayed on their website. In the last US Presidential campaign, some false news stories, such as the one in which the Pope endorses Donald Trump, became viral and brought substantial traffic to these websites. Over 100 of these websites were located in Macedonia (Silverman & Alexander, 2016).

**Disinformation Tactics to Erode Trust in Democratic Institutions**

The first strategic aim as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) concerns affecting the link between political figures and citizens, and more precisely how citizens perceive political leadership. The intent is to increase public support for a political party or a political leader.

These strategies are executed through two main tactics: false amplification (bots and fake accounts) and imposter content (Wardle, 2017) (astroturfing, impersonation). They include the amplification or fabrication of critiques (Brown, Parrish, & Speri, 2017) and trends (including conspiracy theories), the impersonation of public (Harding, 2018) and political figures (Kelly, Truong, Shahbaz, Earp, & White, 2017) and political opponents (Reporters Without Borders, 2018), and faking support from grassroots organizations (astroturfing) (Kelly, Truong, Shahbaz, Earp, & White, 2017) to show support from different sources. Although part of the support may be artificially created, these measures can also trigger real support if the audience does not realize that part of the support is fabricated.

The second strategic aim of disinformation campaigns as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) regards political participation, and more particularly reducing the participation of citizens in electoral processes in order to support political opponents. This aim is quite

broad since it includes strategies to undermine trust in democracy and electoral processes, foster polarization, and suppress voices.

In terms of tactics, it includes interfering with the political process, concentrating on/utilizing harassment and intimidation to keep some voices out of the information spaces (defamation, e.g. white trolls against journalists in Turkey; Kelly, Truong, Shahbaz, Earp, & White, 2017), and exploiting legislative systems that are not up-to-date with the large array of uses of digital technologies (dark advertising).

Two narratives support this first grand strategy of global leadership. The first one showcases the authoritarian regimes' successful response to Covid-19, and the second one depicts how weak European pluralist democracies are.

The first narrative highlighted the successful management of the crisis by China, Turkey, and Iran. CRI in German depicted China as the leading force that drove the global economic recovery (CRI, 2020d). TRT in Spanish showcased the Turkish healthcare system as one of the best in the world to combat Covid-19 (TRT, 2020b), and alleged its international collaboration in the development of a new vaccine (TRT, 2020d), including a new radiation system being tested in the US (TRT, 2020e). HispanTV showed Iran's support provided to Kyrgyzstan (HispanTV, 2020a) and praised Cuban doctors for allegedly taking care of over 26,000 Covid-19 patients in the world (HispanTV, 2020b). RT in German reported on an Italian businessman who changed the EU flag for the Russian one "to thank Russia for sending 12 planes [while] the EU closes everything down" (RT, 2020a).

The second narrative focused on the weak response of Western democracies to Covid-19. Sputnik and RT in French and German mentioned stories about the lockdown and civil unrest in France (RT, 2020b), Italy (RT, 2020g), Germany (Sputnik, 2020a; RT, 2020f), and Poland (RT, 2020d), including the violent protests involving the yellow vests (RT, 2020e). They also framed some "real stories" differently, including how healthcare workers in Belgium turned their backs on the Prime Minister who arrived at their hospital (RT, 2020b, 2020c), and the increase in the gap between the rich and poor in Germany (RT, 2020h). TRT in English presented homeless people in Europe (TRT, 2020a), and highlighted the difficult situation of refugees living in camps in Greece (TRT, 2020c). TRT also depicted France as a country where "discrimination is rampant" and constitutes a "societal sickness" (Ramadani, 2020). RT in English highlighted the fact that one of the main pillars of European liberalism – free movement of people – was being torn apart (Dockery, 2020).

Chinese media intended to undermine the credibility of US political leadership (CRI, 2020a, 2020e). Xinhua in French claimed that the US were diffusing a political virus (Xinhua News Agency, 2020b), and CGTN in French argued that the US staged a "Hollywood-style" show to distract from its disastrous Covid-19 management (CGTN, 2020a). CRI in German and

Spanish stated that US leadership was racist and self-serving and was bringing their country to a fall (CRI, 2020b, 2020c). CRI in German further highlighted how income inequalities among the US population (CRI, 2020g) and alleged incompetence of the US political leadership exemplified the failure of the US democratic model (CRI, 2020f, 2020g). HispanTV alleged that the Russian foreign minister initiated talks with other state representatives to clarify U.S. military and microbial activities in various regions, including near Russia's borders (HispanTV, 2020c).

**Disinformation Tactics to Erode Trust in the News Ecosystem**

The third strategic aim of disinformation campaigns as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) is to target the integrity of the communications environment itself: when compromised, anti-government protests cannot be coordinated and discourses are undermined (Shiffrin, 2014). The aim is to disrupt communication channels and create a digital environment that citizens no longer trust. In terms of tactics, it includes the abuse of content moderation, playing both sides to foster anger and confusion, fabricating and diffusing scare stories, shocking or graphic content, and dominating online discourse with hashtag poisoning and spam diffusion (Kelly, Truong, Shahbaz, Earp, & White, 2017).

The last strategic aim of disinformation campaigns as identified by Krasodomski-Jones, Smith, Jones, Judson, and Miller (2019) relates to the quality of the information citizens can access; more precisely the aim is to create information chaos, where it is not clear any longer what is true and what is false. Consequently, facts lose their value, and debate reaches an epistemic paralysis, a post-truth, or weaponized relativism (*The Guardian*, 2015). The strategies associated with this aim focus on undermining trust in media and digital media, and affecting the content produced (fabricating, suppressing, or promoting content).

To support this broad aim, a large array of tactics is available, including false news as will be discussed in the next subsection. In addition, tactics also include the exploitation and manipulation of algorithms, suppressing access to some content, and diffusing conspiracy theories. According to Nisbet and Kamenchuk (2019), another tactic is called information gaslighting. Gaslighting comes from the British play entitled *Gas Light* and its later 1940 and 1944 film adaptations, where systematically a husband psychologically manipulates his wife. In the context of disinformation, information gaslighting describes the fast production of false and contradictory information. As Adkins (2019) argues, citizens can no longer distinguish between reality and fantasy

when exposed to the accumulation of disinformation. In other words, information gaslighting alters the target's perception of reality:

> Gaslighting exploits weaknesses in the human mind and has a debilitating effect on the victim's ability to think rationally and to function independently of the gaslighter. It can take many forms. In all instances, however, it involves the clever manipulation of "reality" by a predator that undermines the victim's independent mental functioning for the gaslighter's own political, financial, or psychological motives. (Welch, 2008, p.6)

Information gaslighting is this form of disinformation campaign that targets the perception capacity of citizens: in a situation of chaos, where no one knows what is correct and what is incorrect, citizens are more easily manipulatable and vulnerable to false news. Information gaslighting confounds citizens and distracts them from what is really happening offline (Nisbet & Kamenchuk, 2019).

Another tactic used during disinformation campaigns, identified by Nisbet and Kamenchuk (2019), has to do with the incidental exposure of citizens to the large variety of forms of false news. It refers to the fact that users of social media platforms can be exposed "by accident" to false news, even when they are not the main target, for instance when false news is discussed and disseminated within a network. This is particularly prevalent on social media platforms, where echo chambers do not favor the exposure to other sources of information and pluralist views, which could contradict the false information disseminated.

Two narratives support the second grand strategy: sowing confusion. This strategy aims at diffusing conspiracy theories about Covid-19, in particular about its origin and about some of the remedies. Disinformation efforts first diffused content to cast doubt on the origin of the virus, intending to make citizens in the world doubt the official version of EU authorities and EU Member States, and European press. For instance, CRI in Spanish showed cases of Covid-19 that could not have their origin in Wuhan, such as a New Jersey mayor who was supposedly infected a long time before the epidemic started in China.

Second, the disinformation efforts intended to make citizens believe that the virus was of military origin and coming from US military bases. Iran's Press TV claimed that the virus came from a "biowarfare" lab based in the US, and HispanTV claimed that the virus came from a US laboratory. CGTN stated in an editorial that the US military may have brought the coronavirus to Wuhan (Fuhua, 2020). CRI in German called for the US to provide an explanation about its biological laboratories in the world with military purposes (CRI, 2020a), while CGTN in Spanish wondered if the US 200+ military biolog-

ical laboratories were preparing new biological weapons and lethal viruses (CGTN, 2020c).

Sputnik in German hinted that the US was leading threatening experiments in epidemiology in Georgia, Kazakhstan, and Armenia and reported on bloggers who claimed that the US was testing bioweapons (Sputnik, 2020b). RT and Sputnik in English diffused content produced by other outlets, in particular Iranian outlets (Nimmo et al., 2020a). Russia disinformation efforts also engaged with other conspiracy theories to gain more engagement. For instance, RT in German cited an Italian politician who had asked for the arrest of Bill Gates for crimes against humanity for the role of the Bill and Melinda Gates Foundation during the Covid-19 pandemic (RT, 2020g). Fact-checkers identified a large volume of conspiracy theories distributed in the Western Balkans about the "man-made" characteristics of the virus and some "miracle cures."

As discussed, disinformation campaigns should be considered part of a grand strategy to challenge the common understanding of the benefits, relevance, and resilience of European liberal democracies, and in doing so, contribute to a global geopolitical power play. The following subsection will present some disinformation operations.

## DISINFORMATION OPERATIONS

This section focuses on disinformation operations led by two countries active in disinformation activities in Europe: China and Russia. Although they are not the only foreign sources of disinformation in Europe, these two countries have a long history of information control both domestically and internationally, which provides scholarship a larger array of data to illustrate disinformation operations. In addition, disinformation not only comes from outside Europe but also from inside as illustrated below, as is the case with some far-right movements. This section illustrates disinformation in Europe with concrete examples[2] of its use to weaken EU institutions and its Member States.

### Changing Narratives during Covid-19

In December 2019, Wuhan, the provincial capital of Hubei in the People's Republic of China became the focus of a new virus: coronavirus. But it was not until 31 December that the authorities in this city published an official statement about the virus, and only in January officially admitted that the virus was transmissible to humans. The authorities placed the whole area under quarantine with the confinement of nearly 60 million people and the building of temporary hospitals to meet the needs of a massive number of sick people.

Chinese President Xi Jinping assured the population that the authorities were doing everything they could to contain the virus.

Among the first victims was a doctor (Li Wenliang) who died on 7 February 2020. He was a whistle blower and alerted the authorities to the virus. Many of his patients were infected with Covid-19. Other doctors sounded the alarm at the end of December. They circulated information in private groups on WeChat, one of the Chinese social media platforms. This equivalent of WhatsApp in China is highly controlled and monitored by the Chinese central authorities. Hence, if they circulated this information in private groups, these doctors probably knew that they could be spied on (Sautedet, 2020). When this information began to filter out, Doctor Li Wenliang was arrested by the local authorities for having, according to Beijing, spread illegal information. His death caused outrage on Chinese social media platforms (Bondaz, 2020) where millions of citizens expressed their anger and did not hesitate to denounce the false information that the Chinese government communicated. Several videos also show Chinese citizens in front of windows shouting "it is all fake" (L'OBS, 2020).

The first reaction of the Chinese authorities was to allow the explosion of anger on social media platforms. Until the end of the month of January, there seemed to be a form of tolerance in the Chinese propaganda apparatus toward news items that were quite critical of the authorities' Covid-19 crisis response (Bondaz, 2020).

However, at the beginning of February, Chinese authorities changed their approach and again started controlling the flow of information very closely. They decided to change the narrative of what happened and their crisis management response. This was done through increased censorship in two stages. The first step of this effort was to eliminate any information that negatively portrayed the role of the public authorities. The second step was to produce and diffuse new content that depicted a positive government response to the crisis, such as the construction of hospitals in just a few days. Doctor Li Wenliang, who had been accused of fomenting a state conspiracy and harming social stability, was now recognized by the same public authorities as a hero (Sautedet, 2020). He was presented as a member of the Communist Party and a martyr. He was being used by the central authorities to stage this denial of responsibility (Bondaz, 2020).

Moreover, disinformation campaigns addressed the origin of the virus. On social media platforms, rumors about the virus coming from outside China were not eliminated, while others about the virus coming from experimental laboratories in Wuhan were systematically eliminated. Starting in mid-February, the Chinese authorities expanded their disinformation campaign to other countries, including Europe. New stories appearing in Chinese media outlets such as Global Times, and relayed by Chinese diplomats abroad, aligned with

the argument that the origin of the virus was not China (Sautedet, 2020). For example, the Chinese media Global Times implied that cases of Covid-19 had been identified in Italy as early as November, referring to a doctor, who in fact had spoken of atypical pneumopathy. He confirmed that he was opposed to this rumor and that these cases had nothing to do with Covid-19 (Bondaz, 2020). Another example of content that was spread abroad through social media platforms was the attempt by the spokesman of the Chinese Ministry of Foreign Affairs to relay rumors about an American origin of the virus, using as an argument that American soldiers had taken part in the military games in Wuhan in autumn 2019 and brought the virus to China. The objective of this information gaslighting form of disinformation campaign was clear: the Chinese authorities were trying to sow doubt and minimize their responsibility.

In this narrative abroad, the central Chinese authorities also included China's health diplomacy, with a message to developing countries: China is an essential partner and this partner is capable of helping them when Europeans and Americans are unable to do so. In Europe, the mask diplomacy is part of the effort to replace the old narrative of mis-management of the crisis with capability and support. Images and declarations about China supplying protective equipment and detection kits flooded social media platforms. Chinese media showcased the support of China to European countries affected by Covid-19, such as the distribution of masks and respirators to Italy (Xinhua News Agency, 2020a), Spain (CGTN, 2020c), and the UK. CGTN published an article entitled "China announces to help 82 countries fight COVID-19" (CGTN, 2020b) highlighting the wide scope of China's international aid. The Chinese *People's Daily* newspaper (2020) celebrated "academicians from the Chinese Academy of Sciences and Chinese Academy of Engineering [who] have become known as 'warriors in white'."

The aim of this disinformation campaign was on the one hand to influence international perceptions of China, and on the other hand to challenge the social imaginary of European liberal democracies. The Chinese authorities deleted content that negatively depicts their management of the crisis, and produced new content aligned with the narrative that China is handling Covid-19 better than the Western democracies. The Covid-19 crisis and its response were an opportunity to show the successes of the Chinese political system, and the superiority of the Chinese system over European liberal democratic systems. In doing so, they hamper the social imaginary of European liberal democracies. A similar objective is pursued by Russian disinformation campaigns.

**Russian "Secondary Infektion" Disinformation Campaign**

The Oxford University computational propaganda project already back in 2017 identified Russian interference in electoral processes in different European

countries, including the United Kingdom, Germany, the Netherlands, Norway, and Sweden. The Russian hacker group called "Fancy Bear," the same one that leaked the emails from the Clinton campaign in 2016, conducted cyberattacks against political leaders and governmental agencies in Germany, while other hackers associated with Russian intelligence gained access to data from the Bundestag.

Russian disinformation campaigns operated through troll factories, hackers, and bots to sow chaos and exploit the vulnerabilities of elections and the public sphere in democratic states (Pomerantsev, 2014) where information flows with few public gatekeepers and many communication channels. In 2013, with an annual budget of USD 10 million, the well-known Internet Research Agency in St. Petersburg, Russia, employed about 600 people (Bugorkova, 2015). Operators were assigned a specific audience, and goals with a precise number of followers to attract. For instance, they were expected to manage at least 10 Twitter handles and tweet 50 times a day on each; manage six Facebook accounts and publish at least three posts daily; and post around 50 articles per day (Bugorkova, 2015). The US Special Prosecutor Robert Mueller further documented the extent of these operations in 2018.

The resulting confusion makes the efforts of states more difficult, since they not only have to ascertain/understand/identify the reality of the field, but then share that reality with the public and businesses. In other words, disinformation campaigns both change what individuals see of reality, and they obstruct the view of states. This change is reflected in the Russian military's understanding of the emergence of a "new generation of warfare" (*voina novogo pokoleniya*), and is well-illustrated by the use of information during the Russian military annexation of Crimea and Ukraine (Thornton, 2015).

Creating confusion is the first tactic of disinformation operators, whether it is in times of crisis or in areas of conflict. By spreading confusion, the state at the origin of the disinformation campaign increases the level of risk for its opponents. Another tactic used by Russia is to alter its image abroad. The Russian narrative of unpredictable leadership is a key element of Russia's disinformation campaigns, as it feeds three other objectives. First, it triggers uncertainty about the real situation on the ground as well as Russia's intentions. Second, it supports dissension within and among other states. Third, it contributes to the perception of a strong Russia (Thornton, 2015).

During the Paris terrorist attacks for instance, governments took time to sort through the false alerts and the real call for help and description of what was happening. The Paris crisis exhibited a large base of social media users tweeting and posting about the attack in a concentrated metropolitan area. This produced an "information cascade," whereby platforms such as Twitter and Facebook were inundated with posts of dubious credibility, thus complicating action on the ground (Melissen & Caesar-Gordon, 2016).

In Ukraine, contradictory information about movements of Russian troops near the eastern border of Ukraine was published before and during the conflict. This effort resulted in buying time in the initial stages of the conflict by thickening the fog of war (Wirtz, 2015). The Russian government also supported bloggers and individuals to broadcast pro-Russian narratives on social media networks (Dougherty, 2014) and sometimes simulate anti-Russian news sources to disseminate false information about the ongoing conflict. "Foreign politicians talk about Russia's interference in elections and referendums around the world. In fact, the matter is even more serious: Russia interferes in your brains, we change your conscience, and there is nothing you can do about it" (Vladislav Surkov, Adviser to Russian president Vladimir Putin, in Maza, 2019).

Information as a tool of power and control has a long history in the world, and particularly in Russia, where it is a systemic phenomenon: it has become part of Russian strategic culture (Darczewska, 2015). Russia and prior to that the USSR have long engaged in information control, manipulation and weaponization of information as discussed previously. Recent and well-documented examples include disinformation in the form of fabricated content about the plane crash MH17 in Ukraine, the 2016 EU–Ukraine Association Agreement referendum, the Crimea annexation, and more broadly the Russia–Ukraine conflict. Other well-known cases include the 2016 Brexit, 2017 French presidential election (DreuxVachon, 2017), and the 2017 Catalonian independence referendum. However, other smaller European countries were also affected in recent years by disinformation stemming from Russia, including Sweden (Nimmo et al., 2020b) and Czechia (Syrovátka, 2019). Russian disinformation efforts are reported to have produced 2,500 pieces of content in seven languages over 300 online platforms since 2014 (Nimmo et al., 2020b). In 2019 Facebook announced "16 accounts, four pages, and one Instagram account as part of a small network emanating from Russia" (Gleicher, 2019b). But the operation was in fact much larger and part of Moscow's decades-long strategic engagement to sow chaos and weaken Western democracies (The Associated Press, 2020).

To illustrate, Yevgeny Primakov, former Director of the Russian Foreign Intelligence Service, recognized that its services supported the diffusion of the narrative claiming that the US Government created the AIDS virus in the 1980s (Kello, 2017). This disinformation campaign was given the name "Operation InfeKtion" by historian Thomas Boghardt (2009), although the real code name was in fact identified later as Operation "DENVER" (Selvage, 2019). The objective of this campaign was to undercut the United States' credibility, promote anti-Americanism, and generate friction between the US and its allies. The narrative about AIDS supported the view that US military bases were the origin of the spread of the virus abroad (US Department of State,

1987). It started with an anonymous letter sent to an Indian journal supporting this scientific claim, and was followed in 1985 by broader diffusion with the help of allied secret service agencies such as the Bulgarian Committee for State Security:

> We are conducting a series of [active] measures in connection with the appearance in recent years in the USA of a new and dangerous disease, "Acquired Immune Deficiency Syndrome – AIDS"…, and its subsequent, large-scale spread to other countries, including those in Western Europe. The goal of these measures is to create a favorable opinion for us abroad that this disease is the result of secret experiments with a new type of biological weapon by the secret services of the USA and the Pentagon that spun out of control. (KGB, 1985)

A false scientific report was then diffused at the summit meeting of the Non-Aligned Movement in 1986 entitled, "AIDS: USA home-made evil, NOT out of AFRICA" (Selvage and Nehring, 2019).

But this narrative and its pseudo-scientific claims were soon denounced by Western and Soviet virologists, and by the Western press, further to letters sent to newspaper editors and journalists by US Embassy officials. Their argument was that it was not scientifically possible at the time to create such a complex virus (US Department of State, 1987). As a result of this international denunciation, the Russian authorities decided to abandon this narrative in 1987 (Andrew & Mitrokhin, 2005).

In 2019, after Facebook deleted a large number of accounts associated with Russian disinformation efforts, researchers at Atlantic Council's Digital Forensic Research Lab (DFRLab) used the term "Secondary Infektion" to describe a new wave of false news that applied the same strategy to create false content: first create a fake account on a forum to plant a false story, often with the image of a counterfeit document to support the claim. Then create a set of fake accounts on various social media platforms to diffuse this story broadly and in different languages. The false news is used in the internet forum as the source to support the claim on other platforms (DFRLab, 2019). What differs from the first Infektion campaign, however, is that the second wave supports several stories.

Contrary to the disinformation strategy adopted during the 2016 US Presidential campaign, the Secondary Infektion aimed to better hide its identity (DFRLab, 2019): By creating a single-use burner account, publishing false news, and then abandoning it to create a new one and publish another false news story or another version of the same story, the disinformation operators not only covered their tracks better, but they also diminished their impact, since they had no time to develop an audience and outreach (The Associated Press, 2020). This second generation of disinformation campaigns continued using platforms such as Reddit, Medium, Twitter, Quora, Facebook, and

YouTube, but also increasingly used blogging forums to diffuse false news and politically explosive stories. It first diffused images of falsely "leaked" documents on blogging forums, and then spread the news on social media platforms (Nimmo et al., 2020b).

In terms of content, the disinformation operators pursued a high drama low impact strategy. Their objective was to generate an emotional response from conspiracy-minded internet communities and make the story viral. One of the stories identified by DFRLab (2019), for instance, claimed that Spanish intelligence unveiled a plot to assassinate Boris Johnson in 2018.

Although the scale of the operation is vast – a large number of channels, it promotes only nine main narratives as identified by the 2020 Graphika report (Nimmo et al., 2020b):

- Russia is the victim (USA and NATO allies are belligerent actors, Turkey is a destabilizing state, Muslims are aggressive invaders, world sporting events are Russophobic, Russia is the victim of Western plots, critics against the Russian government stem from morally corrupt, alcoholic, or mentally unstable individuals);
- Western democracies are weak (Europe is weak and divided, Western democratic elections are rigged);
- Ukraine is unreliable and a failed state.

This second wave of disinformation campaigns did not mainly focus on election interference, as often argued in the press. The objective was more about the traditional geopolitical power strategy: Divide to better conquer. It aimed to intensify divisions between Western countries, such as Poland against Germany, Germany against the USA, Germany against the UK, and everyone against Ukraine (Nimmo et al., 2020b).

Moreover, this second wave of disinformation made intensive use of counterfeit documents to support its claims, including false communications (tweets, letters, and blog posts) from political leaders of Western democracies, such as US Secretary of State Mike Pompeo, representatives of the German, British, and Ukrainian governments, and former national leaders including Carl Bildt (Sweden). The disinformation operators also counterfeited content from nonprofit organizations, ranging from the World Anti-Doping Agency (WADA) and the Organization for Security and Co-operation in Europe (OSCE) to the Committee to Protect Journalists (CPJ) and the environmental group Greenpeace (Nimmo et al., 2020b). Although this tactic is not unique, the volume, consistency, and persistence are. In that sense, it is aligned with previous Russian active measures (Rid, 2020).

This disinformation campaign has unique features, including a very limited impact in terms of engagement metrics when compared to past efforts:

> Almost none of the operation's posts across six years of activity achieved any measurable engagement, in terms of shares, likes and positive reactions across platforms. This may indicate that the operators were not interested in engagement metrics – for example, if they were driven by production quotas rather than engagement targets – or that they were using some other form of metrics not visible to outside observers. The lasting mismatch between effort expended and apparent impact gained is yet another mystery about this operation. (Nimmo et al., 2020b)

In other words, disinformation operations do not necessarily become viral. This means that it is crucial not to inflate their potential impact on political outcomes and polarization. However, their impact is difficult to assess as it can have longer-term and secondary effects, including reduced trust in public institutions, and lack of interest in political representative processes.

Russia's disinformation efforts are now well-documented. However, it is important to avoid inflating its power more than it actually is (Győri & Krekó, 2019). The recent disinformation environment is more complex than before, with tactics and procedures different than the ones developed by the Internet Research Agency and Russia's GRU military intelligence. Moreover, Iran is also now an active disinformation actor on Western social media platforms (Nimmo et al., 2020b), China has become more aggressive abroad (Twitter, 2019), and Western political parties also run their own disinformation campaigns (Gleicher, 2019a).

At the same time, if the actors and techniques of disinformation have evolved, the response has also improved with more researchers involved, tech companies adopting new measures, states adopting new legislations, and citizens more aware of this issue: "The repeated exposure of Secondary Infektion's operations by platforms, journalists, and researchers may have triggered the steep drop in output observed in July 2019 and January 2020. If this model can be continued and reinforced, our collective defenses will be in a significantly better state than in 2016" (Nimmo et al., 2020b).

As discussed in this subsection, EU institutions and European Member States are the targets of large disinformation operations, whose objective is to challenge the common understanding of the benefits, relevance, and resilience of European liberal democracies. In this context, EU institutions and Member States responded gradually and with different means to this threat. The 2019 EU Parliamentary elections were particularly under scrutiny.

> From January to May, online platforms have taken action against inauthentic behaviour to limit the scope of spam and disinformation globally. Google reported to have globally removed more than 3.39 million YouTube channels and 8,600 channels

> for violations against its spam and impersonation policies. Facebook disabled 2.19 billion fake accounts in the first quarter of 2019 and acted specifically against 1,574 non-EU-based and 168 EU-based pages, groups and accounts engaged in inauthentic behaviour targeting EU Member States. Twitter challenged almost 77 million spam-like or fake accounts globally. (EU, 2019, p.4)

As discussed, disinformation campaigns from China and Russia aim to challenge European liberal democracies. By doing so, they aim to hinder how liberal democracies and their institutions are perceived by their citizens and other countries in the world. Hence, disinformation campaigns are part of a global power play to reduce the influence and the role of liberal democracies and democratic values in the world. The following section presents the main efforts of EU institutions to combat disinformation campaigns in this context.

## RESPONSE TO DISINFORMATION CAMPAIGNS

With the growth of false news and disinformation campaigns, fact-checking has become one of the objectives of online platforms, the press, and Western governments. ReportersLab identified about 160 fact-checking organizations in the world (Lim, 2019). In Europe, some organizations such as EUFactcheck.eu[3] or EUvsDisinfo.eu[4] are prime examples of the efforts of the press and EU institutions to combat the spread of false news. This point is discussed in more detail in a subsequent chapter of this book. Facing growing disapproval from their users, and a new set of regulations to force them to take down false content, online platforms turned to AI to increase their content moderation capacity and as much as possible automate it. Part of the moderation is still done by human operators based in developing countries. Automated fact-checking (AFC) technologies pursue three objectives: "to spot false or questionable claims circulating online and in other media; to authoritatively verify claims or stories that are in doubt, or to facilitate their verification by journalists and members of the public; and to deliver corrections instantaneously, across different media" (Graves, 2018, p.2). Although promising, this technology faces several challenges, since fact-checking requires judgment and sensitivity to context, which fully automated fact-checking systems cannot do. AFC are particularly challenged by conversational sources, such as discussion on social media platforms, where users use pronouns and refer back to earlier points, or use words with double meanings. So far, AFC technologies prove useful predominantly "to assist fact-checkers to identify and investigate claims, and to deliver their conclusions as effectively as possible" (Graves, 2018, p.1).

In 2015, the European Council "stressed the need to challenge Russia's ongoing disinformation campaigns and invited the High Representative, in cooperation with Member States and EU institutions, to prepare by June an

action plan on strategic communication. The establishment of a communication team is a first step in this regard." In response, the East Stratcom Task Force[5] was created as part of the European External Action Service (EEAS). This task force has identified and catalogued over 9,000 examples of pro-Kremlin disinformation. The website EUvsDISINFO that the East Stratcom Task Force then created raises awareness about these "messages in the international information space that are identified as providing a partial, distorted, or false depiction of reality and spread key pro-Kremlin messages."

The years 2016 and 2017 marked a change in the international context. Former UK Prime Minister Theresa May claimed that Russia was "weaponizing information" and the US acted on Russian interference in the US Presidential election by imposing a number of sanctions against diplomats and individuals associated with the Internet Research Agency. The 2019 Report by Special Counsel Robert Mueller stated that "The Russian government interfered in the 2016 presidential election in sweeping and systematic fashion. Evidence of Russian government operations began to surface in mid-2016." Further, "The campaign evolved from a generalized program designed in 2014 and 2015 to undermine the U.S. electoral system, to a targeted operation that by early 2016 favored candidate Trump and disparaged candidate Clinton" (Mueller, 2019).

In 2016, the Joint Framework on countering hybrid threats "encouraged a whole-of-government approach, with 22 areas for action, to help counter hybrid threats and foster the resilience of the EU and the Member States," and recommended a series of actions "ranging from bolstering EU's intelligence analysis capacity to strengthening protection of critical infrastructure and cybersecurity to fighting radicalisation and violent extremism" (EU, 2016). It also led to the creation of the Hybrid Fusion Cell as the focal point for all EU institutions for the analysis of hybrid threats. It was followed a year later by the establishment in Helsinki of the European Centre of Excellence for Countering Hybrid Threats.[6]

In June 2017, the European Parliament adopted a resolution on "Online platforms and the Digital Single Market" that "Stresses the importance of taking action against the dissemination of fake news" and "Calls on the Commission to analyze in depth the current situation and legal framework with regard to fake news, and to verify the possibility of legislative intervention to limit the dissemination and spreading of fake content" (EU Parliament, 2017). At the end of 2017–beginning of 2018, the EU Commission launched a public consultation on fake news and online disinformation with two questionnaires: "one for the citizens and one for legal persons and journalists reflecting their professional experience of fake news and online disinformation" (EU, 2018e). In a 2018 Eurobarometer survey, 73% of people interviewed expressed their

concern about the impact of false news for the upcoming European Parliament elections (EU, 2018h).

This consultation showed that "Intentional disinformation aimed at influencing elections and immigration policies were the two top categories considered likely to cause harm to society, according to respondents to a public consultation conducted by the Commission" (EU, 2018e).

> A majority of respondents to the public consultation considered that educating and empowering users to better access and use online information and informing users when content is generated or spread by a bot are measures online platforms can take that would have a strong impact on preventing the spread of disinformation. (EU, 2018e)

In January 2018, the year before the EU Parliamentary elections, then EU Commissioner for the Digital Economy and Society Mariya Gabriel convened the High-Level Expert Group (HLEG) to advise the European Commission on disinformation campaigns (2018), and their recommendations were based on five pillars:

1. Augmenting transparency of online news;
2. Encouraging media and information literacy;
3. Developing new tools to empower citizens and journalists to tackle disinformation;
4. Protecting the diversity and sustainability of the news media ecosystem;
5. Promoting continued research on the impact of disinformation (HLEG, 2018).

In April, the EU Commission took on these recommendations and the results of the public consultations to develop a European approach to tackle online disinformation:

> A well-functioning, free, and pluralistic information ecosystem, based on high professional standards, is indispensable to a healthy democratic debate. The Commission is attentive to the threats posed by disinformation for our open and democratic societies. This Communication presents a comprehensive approach that aims at responding to those serious threats by promoting digital ecosystems based on transparency and privileging high-quality information, empowering citizens against disinformation, and protecting our democracies and policy-making processes. (EU, 2018e)

The Communication recommended the creation of an EU-wide independent network of fact-checkers and initiatives to enhance the quality of journalism and augment the digital media literacy of citizens (Bentzen, 2019). This Communication also called for the organization of a Multi-Stakeholder Forum

on Disinformation (2018) to be composed of "representatives of online platforms, the advertising industry and advertisers, as well as academics, media and civil society organisations" (EU, 2018c).

This Forum led to the creation of an EU-wide Code of Practice to improve the explicability of information selection by algorithms and the accessibility to reliable news:

> The Code of Practice on disinformation is the first worldwide self-regulatory set of standards to fight disinformation voluntarily signed by platforms, leading social networks, advertisers and advertising industry in October 2018. Signatories are Facebook, Twitter, Mozilla, Google and associations and members of the advertising industry. Microsoft subscribed to the Code of Practice in May 2019. TikTok joined the code in June 2020. (EU, 2018b)

This EU Code of Practice calls for "deleting fake accounts, labelling messaging activities by 'bots' and cooperating with fact-checkers and researchers to detect disinformation and make fact-checked content more visible" (Bentzen, 2019).

In June 2018, the European Council called on the EU to "protect the Union's democratic systems and combat disinformation, including in the context of the upcoming European election" (EU, 2018c). The same month, the European Commission and the Vice-President of the Commission/High Representative of the Union for Foreign Affairs and Security Policy (HR) wrote a joint communication on boosting resilience against hybrid threats, emphasizing strategic communications as a priority (EU, 2018d).

In September 2018, on the occasion of his State of the Union Address, former EU President Jean-Claude Juncker set out a series of new measures to ensure free and fair 2019 European Parliament elections, including enhanced transparency of online political advertisements, and possible sanctions for the illegal use of personal data (EU, 2018a). The same year, the EU Commission published a Joint Communication on increasing resilience and bolstering capabilities to address hybrid threats in 2018 "to address disinformation emanating from inside and outside the EU and to deter hostile disinformation production and hybrid interference by foreign governments" (EU, 2018d).

In December 2018, the EU Commission developed an "Action Plan against Disinformation" with four pillars:

1. Additional funding, specialized staff, and data analysis tools provided to the Strategic Communication Task Forces, the EU Hybrid Fusion Cell, and the EU delegations in neighborhood countries;
2. Creation of the Rapid Alert System (RAS), a platform where EU Member States and EU institutions can share insights on disinformation in real time and coordinate responses;

3. Call to the online platform companies to effectively implement the commitments they agreed to when signing the EU-wide Code of Practice on Disinformation (signed on 26 September 2018);
4. Promotion of media literacy, creation of targeted awareness campaigns about disinformation, and support to national teams of independent fact-checkers and researchers (Bentzen, 2019).

The Commission launched the Social Observatory for Disinformation and Social Media Analysis (SOMA) facilitating networking, knowledge exchange and development of best practices among independent fact checkers. A first group of 14 European fact-checking organisations have access to SOMA, which is also launching multidisciplinary centres for research on disinformation. The Connecting Europe Facility will also provide funding (EUR 2.5 million) for a new digital service infrastructure aimed at networking fact checkers and researchers (EU, 2019).

As shown, the EU institutions and EU Member State representatives took the threat of disinformation seriously prior to the 2019 elections. The European Commissioner for Security Julian King argued in early 2019 that European elections were "Europe's most hackable election" (Becker, 2019). Although the impacts of disinformation campaigns on electoral outcomes are still debated among experts and scholars (Tucker et al., 2018), their existence is proven by a large array of evidence, as is their purpose to influence public opinion and interfere with the outcome of democratic elections. The study conducted by Avaaz and the Institute for Strategic Dialogue showed that far-right political groups used Facebook to spread disinformation in Germany, the UK, France, Italy, Poland, and Spain prior to the elections (Institute for Strategic Dialogue & Avaaz, 2019).

The 2019 European elections illustrate well the efforts of foreign states to interfere in the democratic processes of other states (Syrovátka, 2019). The cybersecurity company SafeGuard published a study in 2019 that received substantial media attention since it claimed that "half of the citizens of EU Member states have come into contact with disinformation from Russian sources" (Boffey, 2019). The study was based on the analysis of the audience of over 6,000 bots and semi-automated accounts on Twitter, Facebook, and YouTube that were apparently connected to Russian disinformation efforts. However, the list of accounts was never published by SafeGuard, meaning that the attribution was not confirmed by other experts and scholars. Since attribution is the main challenge to assessing the impact of disinformation campaigns, this study must be considered with caution (Syrovátka, 2019).

After the elections, the EU highlighted that "The preliminary analysis shows that it contributed to expose disinformation attempts and to preserve the integrity of the elections, while protecting freedom of expression" and "Malicious

sources, both within and outside the EU, are constantly using new tactics, opting increasingly for smaller-scale local operations that are less likely to be detected and exposed. However, the objective remains the same: dividing our society and undermining the trust of citizens in democratic processes and institutions." (EU, 2019, p.9).

## CONCLUDING REMARKS

This chapter explored contemporary forms of disinformation, focusing on disinformation campaigns through social media platforms in Europe. AI is both a defense against and a weapon to support disinformation campaigns. On the one hand, AI is offered as the main solution of filtering out false news. On the other hand, AI is used to spread false news either through the MLA of social media platforms (Hao, 2021) or automated tools such as bots.

Disinformation campaigns target the trust established between citizen governments (Barela & Duberry, 2021), as well as their trust in the information ecosystem:

> [d]isinformation erodes trust in institutions and in digital and traditional media, and harms our democracies by hampering the ability of citizens to take informed decisions. Disinformation also often supports radical and extremist ideas and activities. It impairs freedom of expression, a fundamental right enshrined in the Charter of Fundamental Rights of the European Union (Charter). (EU, 2018f, p.1)

Disinformation campaigns aim to hinder how liberal democracies and their institutions are perceived by their citizens and other countries in the world. Hence, disinformation campaigns are part of a global power play to reduce the influence and role of liberal democracies and democratic values in the world. Disinformation campaigns must also be understood from this global perspective, where authoritarian regimes push the narrative on social media platforms that liberal democracies are not relevant and resilient, and focus on the challenges they present rather than on the benefits they offer to their citizens. AI is at the center of this battlefield both as an enabler of disinformation diffusion by controlling content distribution (i.e. MLA of online platforms favoring juicy content), and as a potential opportunity to mitigate their diffusion (i.e. automated fact-checking).

## NOTES

1. Two memos summarize their latest findings at the time of writing this book: "Coronavirus Coverage by State-Backed English-Language News Sources: Understanding Chinese, Iranian, Russian and Turkish Government Media." Data Memo 2020.2; and "Covid-19 News and Information from State-Backed Outlets

Targeting French, German and Spanish-Speaking Social Media Users." Data Memo 2020.3. comprop.oii.ox.ac.uk [Accessed 22 September 2021].
2. The examples are not exhaustive but provide a real-life illustration.
3. See website EUFactcheck.eu. https://eufactcheck.eu/about-us/ [Accessed 22 September 2021].
4. See website EUvsDisinfo. https://euvsdisinfo.eu [Accessed 22 September 2021].
5. See website Stratcom. https://stratcomcoe.org [Accessed 22 September 2021].
6. See website Hybridcoe. https://www.hybridcoe.fi [Accessed 22 September 2021].

## REFERENCES

Adamsky, D. (2015). Cross-domain coercion: The current Russian art of strategy. *Proliferation Papers*, 54 (November), 1–43.

Adkins, K. C. (2019). Gaslighting by crowd. *Social Philosophy Today*, 35, 75–87.

Andrew, C. & Mitrokhin, V. (2005). *The World Was Going Our Way: The KGB and the Battle for the Third World: Newly Revealed Secrets from the Mitrokhin Archive*. New York: Basic Books.

Arquilla, J. & Ronfeldt, D. (1993). Cyberwar is coming! *Comparative Strategy*, 12(2), 141–165.

Barela, S. J. & Duberry, J. (2021). Understanding disinformation operations in the 21st century. In: Hollis, D. B. and Ohlin, J. D. (eds.), *Defending Democracies: Combating Foreign Election Interference in a Digital Age*. Oxford: Oxford University Press, pp.41–72.

Becker, I. (2019). Europe's most hackable election. *Encompass*, 16 January. https://encompass-europe.com/comment/europes-most-hackable-election-threats-of-electoral-manipulation-in-the-digital-age [Accessed 22 September 2021].

Bennett, W. L. & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, 33(2), 122–139.

Bentzen, N. (2019). Online disinformation and the EU's response. *European Parliament Research Service*. https://www.europarl.europa.eu/RegData/etudes/ATAG/2018/620230/EPRS_ATA(2018)620230_EN.pdf [Accessed 22 September 2021].

Bittman, L. (1981). Soviet Bloc "disinformation" and other "active measures". In: Pfaltzgraff, R. L., Ra'anan, U., and Milberg, W. (eds.), *Intelligence Policy and National Security*. London: Palgrave Macmillan, pp.212–228.

Boffey, D. (2019). 241m Europeans "may have received Russian-linked disinformation". *The Guardian*, 8 May. https://www.theguardian.com/world/2019/may/08/241m-europeans-may-have-received-russian-linked-disinformation [Accessed 22 September 2021].

Boghardt, T. (2009). Soviet Bloc intelligence and its AIDS disinformation campaign (Operation INFEKTION). *Studies in Intelligence*, 53(4), 1–24.

Bondaz, A. (2020). Covid-19, une épidémie de fausses nouvelles. *Mecanisme du Complotisme. Podcast France Culture*. https://www.franceculture.fr/emissions/mecaniques-du-complotisme [Accessed 22 September 2021].

Bradshaw, S. & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Project on Computational Propaganda.

Brown, A., Parrish, W., & Speri, A. (2017). Leaked documents reveal counter-terrorism tactics used at Standing Rock to defeat pipeline insurgencies. *The Intercept*. https://theintercept.com/2017/05/27/leaked-documents-reveal-security-firms-counterterrorism-tactics-at-standing-rock-to-defeat-pipeline-insurgencies/ [Accessed 22 September 2021].

Bugorkova, O. (2015). Ukraine conflict: Inside Russia's "Kremlin troll army". *BBC*. https://www.bbc.com/news/world-europe-31962644 [Accessed 22 September 2021].

Burenok, V. M. (2018). *Kontseptsii perspektivnogo oblika silovykh kompontentov voennoy organizatsii Rossiyskoi Federatsii* [Concepts of the Perspective Appearance of the Power Components of the Military Organization of the Russian Federation]. Moscow: Russian Academy of Missile and Artillery Sciences (RARAN).

Cavelty, M. D., Mauer, V. (2008) The role of the state in securing the information age— challenges and prospects. In: Cavelty, M. D., Mauer, V., Krishna-Hensel, S .F. (eds.), *Power and Security in the Information Age*, Ashgate Publishing.

CGTN. (2020a, 21 May). Commentaire: Les Politiciens Américains, Auteurs Des Mensonges Inscrits Dans L'histoire [Commentary: American Politicians, Authors of Lies Written Into History]. *CGTN*. https://francais.cgtn.com/n/bfjaa-bea-fea/cbeiia/index.html [Accessed 22 September 2021].

CGTN. (2020b, 21 March). China announces to help 82 countries fight COVID-19. *CGTN*. https://news.cgtn.com/news/2020-03-20/China-announces-to-help-82-countries-fight-COVID-19-P1hcQcQKe4/index.html [Accessed 22 September 2021].

CGTN. (2020c, 19 May). Búsqueda de Emergencia Global: ¿Dónde Están Los Más De 200 Laboratorios Biológicos En EE. UU? https://espanol.cgtn.com/n/2020-05-19/djcdea/busqueda-de-emergencia-global-donde-estan-los-mas-de-200-laboratorios-biologicos-en-eeuu/index.html [Accessed 22 September 2021].

Chekinov, S. & Bogdanov, S. (2015a). Voennoe iskusstvo na nachal'nom etape XXI stoletiya: problemy i suzhdeniia [Military art in the initial stage of the XXI century: problems and judgments]. *Voennaya Mysl*', no. 1 (January), 34–45.

Chekinov, S. & Bogdanov, S. (2015b). Prognozirovanie kharaktera i soderzhaniia voin budushchego: problemy i suzhdeniia [Predicting the character and content of a future warrior: challenges and judgments]. *Voennaya Mysl'*, no. 10 (October), 41–49.

Chivvis, C. (2017). Hybrid war: Russian contemporary political warfare. *Bulletin of the Atomic Scientists*, (August), 316–321.

Clausewitz, C. (2003). *On War*. London: Penguin Random House.

CNN. (n.d.). https://edition.cnn.com/2020/02/06/politics/senate-intel-report-russian-interference-obama-response/index.html [Accessed 22 September 2021].

CRI. (2020a, 17 May). Die USA Schulden Der Welt Eine Erklärung Für Ihre Biologischen Labors In Übersee. http://german.cri.cn/aktuell/alle/3250/20200517/468493.html [Accessed 22 September 2021].

CRI. (2020b, 18 May). Qué Teatro Hacen Esos Políticos Estadounidenses Con Mentiras Una Tras Otra? Español. http://espanol.cri.cn/news/report/1017/20200518/468820.html [Accessed 22 September 2021].

CRI. (2020c, 18 May). Rassistische US-Politiker Müssen Niederlage Schlucken. http://german.cri.cn/kommentar/alle/3259/20200518/469024.html [Accessed 22 September 2021].

CRI. (2020d, 19 May). Chinas Einsatz Belebt Globale Zuversicht In Pandemiebekämpfung In Kritischem Moment. http://german.cri.cn/kommentar/alle/3259/20200519/469444.html [Accessed 22 September 2021].

CRI. (2020e, 19 May). Untersuchung Des US-Kongresses: 500 Milliarden-Rettungspaket Liegt Brach. http://german.cri.cn/aktuell/alle/3250/20200519/469317.html [Accessed 22 September 2021].

CRI. (2020f, 22 May). Verfälschung Von Krankenund Todesdaten Verdeutlicht Versagen Von Uspandemieprävention. http://german.cri.cn/kommentar/alle/3259/20200522/471617.html [Accessed 22 September 2021].

CRI. (2020g, 23 May). Covid-19-Pandemie Enthüllt Die Wahrheit Der Großen Ungleichheit In Den USA. http://german.cri.cn/kommentar/alle/3259/20200523/472019.html [Accessed 22 September 2021].

Darczewska, J. (2015). The devil is in the details. Information warfare in the light of Russia's military doctrine, point of view. *Centre for Eastern Studies*, May 2015. http://goo.gl/UUrYux, 9 [Accessed 22 September 2021].

Deibert, R. J. & Pauly, L. W. (2019). Mutual entanglement and complex sovereignty in cyberspace. In: Bigo, D., Isin, E., and Ruppert, E. (eds.), *Data Politics*. London: Routledge, pp.81–99.

DFRLab. (2019). Top takes: Suspected Russian intelligence operation. *Medium*. https://medium.com/dfrlab/top-takes-suspected-russian-intelligence-operation-39212367d2f0 [Accessed 22 September 2021].

Dockery, G. (2020). Covid-19 dismantles the hollow commandments of European liberalism. *RT*, 20 March 2020.

Dougherty, J. (2014). Everyone lies: The Ukraine conflict and Russia's media transformation. *Shorenstein Center on Media, Politics and Public Policy Discussion Paper Series* (July 2014).

DreuxVachon. (2017, 16 March). Emmanuel Macron: A new hope of migrants in Europe? *Medium*. http://web.archive.org/web/20200418160806/https://medium.com/@DreuxVachon/emmanuel-macron-a-new-hope-of-migrants-in-europe-b6d07e0a111b [Accessed 22 September 2021].

EU. (2016). *Joint Framework on Countering Hybrid Threats: A European Union Response*. Luxembourg: EU Publications Office. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016JC0018&from=en [Accessed 22 September 2021].

EU. (2018a). *State of the Union*. https://ec.europa.eu/commission/presscorner/detail/en/IP_18_5681 [Accessed 22 September 2021].

EU. (2018b). *EU Code of Practice on Disinformation*. https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation [Accessed 22 September 2021].

EU. (2018c). *Action plan against disinformation*. Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Publication Office of the European Union https://ec.europa.eu/commission/sites/beta-political/files/eu-communication-disinformation-euco-05122018_en.pdf [Accessed 22 September 2021].

EU. (2018d). *Increasing resilience and bolstering capabilities to address hybrid threats*. Joint Communication to the European Parliament, the European Council and the Council. Publications Office of the European Union, Luxembourg. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018JC0016&from=EN [Accessed 22 September 2021].

EU. (2018e). *Public consultation on fake news and online disinformation*. https://ec.europa.eu/digital-single-market/en/news/public-consultation-fake-news-and-online-disinformation [Accessed 22 September 2021].

EU. (2018f). *Tackling online disinformation: A European approach*. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN [Accessed 22 September 2021].
EU. (2018g). *Action Plan Against Disinformation*. https://eeas.europa.eu/sites/eeas/files/action_plan_against_disinformation.pdf [Accessed 22 September 2021].
EU. (2018h). Special Eurobarometer 477: Democracy and elections. November. https://data.europa.eu/euodp/en/data/dataset/S2198_90_1_477_ENG [Accessed 22 September 2021].
EU. (2019). *Report on the implementation of the Action Plan Against Disinformation.* Joint Communication To The European Parliament, The European Council, The Council, The European Economic And Social Committee And The Committee Of The Regions. JOIN/2019/12 final. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=comnat%3AJOIN_2019_0012_FIN [Accessed 22 March 2022].
European Council. (2015). *EUCO Conclusions*. Press releases 19 and 20 March 2015. https://www.consilium.europa.eu/media/21888/european-council-conclusions-19-20-march-2015-en.pdf [Accessed 22 September 2021].
European Council. (2018). *EUCO Conclusions.* Press Release 28 June 2018. https://www.consilium.europa.eu/en/press/press-releases/2018/06/29/20180628-euco-conclusions-final/ [Accessed 22 September 2021].
European Parliament. (2017). *Online platforms and the Digital Single Market*. European Parliament resolution of 15 June 2017 on online platforms and the digital single market (2016/2276(INI)). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0272_EN.pdf?redirect [Accessed 22 September 2021].
Fuhua, W. (2020). 10 questions for the U.S.: Where did the novel coronavirus come from? *CGTN*, 19 March 2020.
Galeotti, M. (2018). I'm sorry for creating the "Gerasimov Doctrine". *Foreign Policy*, 5 March 2018. https://foreignpolicy.com/2018/03/05/im-sorry-for-creating-the-gerasimov-doctrine/ [Accessed 22 September 2021].
Gartzke, E. (2013). The myth of cyberwar: Bringing war in cyberspace back down to earth. *International Security*, 38(2), 41–73.
Gerasimov, V. (2013). Tsennost' Nauki v Predvidenii [The Value of Science is in Foresight]. *Voyenno Promyshlennyy Kuryer*, 26 February 2013. http://vpk-news.ru/articles/14632 [Accessed 22 September 2021].
Ghernaouti, S. (2013). *Cyber Power: Crime, Conflict and Security in Cyberspace*. Lausanne: EPFL Press.
Gleicher, N. (2019a). *Removing Coordinated Inauthentic Behavior in Spain*. Facebook, 20 September 2019, https://about.fb.com/news/2019/09/removing-coordinated-inauthentic-behavior-in-spain/ [Accessed 22 September 2021].
Gleicher, N. (2019b). *Removing More Coordinated Inauthentic Behavior From Russia*. Facebook. https://about.fb.com/news/2019/05/more-cib-from-russia/ [Accessed 22 September 2021].
Golovchenko, Y., Hartmann, M., & Adler-Nissen, R. (2018). State, media and civil society in the information warfare over Ukraine: Citizen curators of digital disinformation. *International Affairs*, 94(5), 975–994.
Graves, D. (2018). *Understanding the promise and limits of automated fact-checking*. Reuters Institute for the Study of Journalism.
Győri, L. & Krekó, P. (2019). Putin's real shadow. *Visegrad Insight*, 30 May. https://visegradinsight.eu/putins-real-shadow/ [Accessed 22 March 2022].
Hao, K. (2021). How Facebook got addicted to spreading misinformation. *MIT Technological Review*. Available at: https://www.technologyreview.com/2021/

03/11/1020600/facebook-responsible-ai-misinformation/ [accessed 22 September 2021].
Harding, X. (2018). Myanmar military members posed as pop stars to spread violence-inducing fake news on Facebook. *MIC*. https://www.mic.com/articles/191899/myanmar-military-members-fake-news-facebook [Accessed 22 September 2021].
HispanTV. (2020a, 19 May). Irán Entrega Ayuda Humanitaria A Kirguistán Para Combatir Covid-19. https://www.hispantv.com/noticias/diplomacia/466433/iran-envia-ayuda-kirguistan-coronavirus [Accessed 22 September 2021].
HispanTV. (2020b, 20 May). Mé Dicos De Cuba Atienden a Más De 26 Mil Afectados Por Covid-19. https://www.hispantv.com/noticias/cuba/466509/medicos-isla-coronavirus-salvar-vida [Accessed 22 September 2021].
HispanTV. (2020c, 26 May). Rusia Anuncia Consultas En OTSC Sobre Armas Biológicas De EEUU. https://www.hispantv.com/noticias/rusia/467086/armas-biologicas-eeuu-lavrov [Accessed 22 September 2021].
HLEG. (2018). *A multi-dimensional approach to disinformation*. Report of the independent high level group on fake news and online disinformation. Luxembourg: EU Publications Office.
Johnson, R. (2017). The changing character of war: Making strategy in the early twenty-first century. *The RUSI Journal*, 162(1), 6–12.
Joint Publication. (2018a). *Cyberspace Operations*. Washington, DC: U.S. Joint Chiefs of Staff, current as of June 2018. https://irp.fas.org/doddir/dod/jp3_12r.pdf [Accessed 22 September 2021].
Joint Publication. (2018b). *DoD Dictionary of Military and Associated Terms*. 1-02. Washington, DC: U.S. Joint Chiefs of Staff, current as of June 2018 [Accessed 22 September 2021].
Jonsson, O. (2019). *The Russian Understanding of War*. Washington, DC: Georgetown University Press.
Kello, L. (2017). *The Virtual Weapon and International Order*. New Haven, CT: Yale University Press.
Kelly, S., Truong, M., Shahbaz, A., Earp, M., & White, J. (2017). Freedom on the Net 2017. *Freedom House*. https://freedomhouse.org/sites/default/files/2020-02/FOTN_2017_Final_compressed.pdf [Accessed 22 September 2021].
KGB. (1985). Information nr. 2955 [to Bulgarian State security], 7 September 1985, *Wilson Center Digital Archive*. https://digitalarchive.wilsoncenter.org/document/208946 [Accessed 22 September 2021].
Kofman, M. (2016). Russian hybrid warfare and other dark arts. *War on the Rocks*, 11 March 2016. https://warontherocks.com/2016/03/russian-hybrid-warfare-and-other-dark-arts/ [Accessed 22 September 2021].
Krasodomski-Jones, A., Smith, J., Jones, E., Judson, E., & Miller, C. (2019). *Warring Songs: Information Operations in the Digital Age*. London: Demos Center.
L'OBS. (2020). "Tout est faux ! ": une dirigeante chinoise conspuée à Wuhan, foyer de l'épidémie de coronavirus. *YouTube*. https://www.youtube.com/watch?v=QRDclNZlmec [Accessed 22 September 2021].
Libicki, M. C. (2009). *Cyberdeterrence and Cyberwar*. Santa Monica: RAND Corporation.
Lilly, B. & Cheravitch, J. (2020, May). The past, present, and future of Russia's cyber strategy and forces. In: *2020 12th International Conference on Cyber Conflict (CyCon)* (Vol. 1300, pp.129–155). IEEE.

Lim, S. (2019). A better ClaimReview to grow a global fact-check database. ReportersLab. Duke University. https://reporterslab.org/a-better-claimreview-to-grow-a-global-fact-check-database/ [Accessed 22 March 2022].
Lindsay, J. R. (2013). Stuxnet and the limits of cyber warfare. *Security Studies*, 22(3), 365–404.
Maza, C. (2019). Vladimir Putin's adviser tells Americans: "Russia interferes in your brains, we change your conscience". *Newsweek*, 12 February 2019. https://www.newsweek.com/rus siapresidentvladimirputinelectionamericans1327793 [Accessed 22 September 2021].
McGuinness, D. (2017). How a cyber attack transformed Estonia. BBC. https://www.bbc.com/news/39655415 [Accessed 22 September 2021].
Melissen, J. & Caesar-Gordon, M. (2016). "Digital diplomacy" and the securing of nationals in a citizen-centric world. *Global Affairs*, 2(3), 321–330.
Ministry of Foreign Affairs of the Russian Federation. (2011). *Konventsiya ob obespechenii mezhdunarodnoy informatsionnoy bezopasnosti (kontseptsiya)* [Convention on International Information Security (Concept)], 22 September. https://www.mid.ru/foreign_policy/official_documents/-/asset_publisher/ CptICkB6BZ29/content/id/191666 [Accessed 22 September 2021].
Molander, R. C., Riddile, A., Wilson, P. A., & Williamson, S. (1996). *Strategic Information Warfare: A New Face of War*. Santa Monica: RAND Corporation.
Morris, L. J., Mazarr, M. J., Hornung, J. W., Pezard, S., Binnendijk, A., & Kepe, M. (2019). *Gaining Competitive Advantage in the Gray Zone: Response Options for Coercive Aggression Below the Threshold of Major War*. Santa Monica: RAND Corporation. https://www.rand.org/ pubs/research_reports/RR2942.html [Accessed 22 September 2021].
Mueller, Robert S. (2019). *Report on the Investigation into Russian Interference in the 2016 Presidential Election*. US Department of Justice.
Multi-Stakeholder Forum on Disinformation. (2018). https://ec.europa.eu/digital-single-market/en/news/meeting-multistakeholder-forum-disinformation [Accessed 22 September 2021].
Nimmo, B. C., Eib, S., Ronzaud, L., Ferreira, R., Lederer, T., & Smith, M. (2020a). Iran's broadcaster: Inauthentic behavior. *Graphika*. https://graphika.com/reports/irans-broadcaster-inauthentic-behavior/ [Accessed 22 September 2021].
Nimmo, B. C., François, C., Eib, S., Ronzaud, L., Ferreira, R., Hernon, C., & Kostelancik, T. (2020b). Exposing secondary infektion. *Graphika*. https://secondaryinfektion.org/downloads/secondary-infektion-report.pdf [Accessed 22 September 2021].
Nisbet, E. C. & Kamenchuk, O. (2019). The psychology of state-sponsored disinformation campaigns and implications for public diplomacy. *The Hague Journal of Diplomacy*, 14(1–2), 65–82.
Nissen, T. E. (2015). *The Weaponization of Social Media – Characteristics of Contemporary Conflicts*. Royal Danish Defence College.
Oberle, J. (2017). Chronique de Jean OBERLE dans "Paris vous parle", 19 janvier 1945. *France Culture*. https://www.franceculture.fr/emissions/concordance-des-temps/lhistoire-vraie-des-fausses-nouvelles-0 [Accessed 22 September 2021].
OECD. (2001). *Engaging Citizens in Policy-Making: Information, Consultation and Public Participation.* Public Management Policy Brief. Paris: OECD Publications.
*People's Daily*. (2020, 19 March). Top Chinese academicians become warriors in battle against novel coronavirus. *People's Daily*. http://en.people.cn/n3/2020/0319/c90000-9669975.html [Accessed 22 September 2021].

Pomerantsev, P. (2014). Russia and the menace of unreality: How Vladimir Putin is revolutionizing information warfare. *The Atlantic*. https://www.theatlantic. com/international/archive/2014/09/russia-putin-revolutionizing-information-warfare/379880/ [Accessed 22 September 2021].
Ramadani, N. (2020). Coronavirus lockdown in France leads to the brutalisation of minorities. *TRT World*, 24 March 2020.
Ramakrishna, K. (2018). Disinformation and fake news: Old wine in new bottles. *RSIS Commentary*, 54.
Reporters Without Borders (2018). *Online Harassment of Journalists: The Trolls Attack*. https://rsf.org/sites/default/files/rsf_report_on_online_harassment.pdf [Accessed 22 September 2021].
Reporters Without Borders (2019). *World Press Freedom Index.* https://rsf.org/en/ranking/2019 [Accessed 22 September 2021].
Rid, T. (2020). *Active Measures: The Secret History of Disinformation and Political Warfare*. London: Profile Books.
RT. (2020a, 20 March). Italienischer Unternehmer ersetzt EU-Fahne durch die Russlands [Italian Entrepreneur Replaces EU Flag with Russia's], *RT*. https://de .rt.com/europa/100059-italienischer-unternehmer-ersetzt-eu-fahne/ [Accessed 22 September 2021].
RT. (2020b, 18 May). Belgique: Les Soignants Réservent Une Haie De Déshonneur À Leur Premier Ministre [Caregivers Give Their Prime Minister a Hail Mary]. *RT*. https://francais.rt.com/international/75150-belgique-soignants-reservent-haie -deshonneur-premier-ministre [Accessed 22 September 2021].
RT. (2020c, 18 May). Kalter Empfang Für Regierungschefin in Belgien: Pflegepersonal Dreht Ihr Bei Besuch Den Rücken Zu [Cold Reception For Head Of Government In Belgium: Nursing Staff Turns Their Backs On Her Visit]. *RT*. https://deutsch.rt.com/kurzclips/102576-kalter-empfang-fur-regierungschefin-krankenhaus/ [Accessed 22 September 2021].
RT. (2020d, 18 May). Polen: Verletzte und Gewaltsame Zusammenstößebei Anti-Lockdown-Protest [Poland: Injured and Violent Clashes during Anti-lockdown Protest]. *RT*. https://deutsch.rt.com/kurzclips/102579-polen-verletzte-und -gewaltsame-zusammenstosse [Accessed 22 September 2021].
RT. (2020e, 18 May). Nach Corona-Lockerungen: Gelbwesten Protestieren Wieder – Festnahmen und Zusammenstöße [After Corona Relaxations: Yellow Vests Protest Again – Arrests and Clashes]. *RT*. Https://Deutsch.Rt.Com/Kurzclips/102587-Nach -Corona-Lockerungen-Gelbwesten-Protestieren/ [Accessed 22 September 2021].
RT. (2020f, 18 May). Unter Dem Medienradar: Brandanschlag Bei Demo Gegen Corona-Beschränkungen In Stuttgart [Under The Media Radar: Arson Attack At Demo Against Corona Restrictions In Stuttgart]. https://Deutsch.Rt.Com/Inland/102577wassonstnochgeschah-Brandschlag/ [Accessed 22 September 2021].
RT. (2020g, 19 May). Italienische Abgeordnete Fordert Verhaftung Von Bill Gates Wegen "Verbrechen An Der Menschlichkeit" [Italian MP Demands Arrest Of Bill Gates For "Crimes Against Humanity"]. *RT*. https://deutsch.rt.com/international/102572-italienische-abgeordnete-fordert-verhaftung-von/ [Accessed 22 September 2021].
RT. (2020h, 19 May). Wenn Die Lobby Fehlt: Bundesregierung Verweigert Corona-Zuschüsse Für Arme [When The Lobby Is Missing: Federal Government Refuses Corona Grants For Poor People]. https://deutsch.rt.com/inland/102622 -bundesregierung-verweigert-zuschusse-fur-arme/ [Accessed 22 September 2021].

Russian Federation Security Council. (2014). *Military Doctrine of the Russian Federation* (approved by the President of the Russian Federation on 25 December 2014), No Pr-2976.

Sautedet, E. (2020). *Covid-19, une épidémie de fausses informations (1/3): complot et fake news made in China*. [Covid-19, an epidemic of false information (1/3): conspiracy and fake news made in China] Podcast Mecanisme du Complotisme. *France Culture.* https://www.franceculture.fr/emissions/mecaniques-du-complotisme [Accessed 22 September 2021].

Schmitt, M. N. (ed.) (2013). *Tallinn Manual on the International Law Applicable to Cyber Warfare*. Cambridge: Cambridge University Press.

Selvage, D. (2019). "Operation Denver": The East German Ministry of State Security and the KGB's AIDS disinformation campaign, 1985–1986 (Part 1). *Journal of Cold War Studies*, 21(4), 71–123.

Selvage, D. & Nehring, C. (2019). Operation "Denver": KGB and Stasi disinformation regarding AIDS. Sources and Methods, A Blog of the History and Public Policy Program. Wilson Centre. https://www.wilsoncenter.org/blog-post/operation-denver-kgb-and-stasi-disinformation-regarding-aids [Accessed 22 March 2022].

Sheldon, J. B. (2011). Deciphering cyberpower: Strategic purpose in peace and war. *Strategic Studies Quarterly*, 5(2), 95–112.

Shiffrin, S. V. (2014). *Speech Matters: On Lying, Morality and the Law*. Princeton: Princeton University Press.

Shil'bakh, K. & Sventsitskiy, V. (1927). *Voennye Razvedki* [Military Intelligence]. Moscow: Military Typography Directorate.

Sidjanski, D. (1979). *Europe Élections de la démocratie européenne*. Paris: Stanké.

Silverman, C. & Alexander, L. (2016). How teens in the Balkans are duping Trump supporters with fake news. *Buzzfeed News*, 3, 874–888.

Spruds, A., Rožukalne, A., Sedlenieks, K., Daugulis, M., Potjomkina, M., Tölgyesi, B., & Bruge, I. (2016). *Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia. Results of the Study*. Latvian Institute of International Affairs Riga Stradins University.

Sputnik. (2020a, 19 May). Nach Angriff Auf Demo-Teilnehmer in Stuttgart: Opfer Schwebt In Lebensgefahr [After Attack On Demo Participants In Stuttgart: Victim's Life Is In Danger]. Sputnik News. https://de.sputniknews.com/gesellschaft/20200519327152209-nach-angriff-auf-demo-teilnehmer-in-stuttgart-opfer-schwebt-in-lebensgefahr/ [accessed 22 September 2021].

Sputnik. (2020b, 29 May). Wozu Benötigen Die USA Biolabore So Nah An Russland? [Why Does The US Need Biolabs So Close To Russia?]. https://de.sputniknews.com/politik/20200529327244206usabiolaborrusslandsnachbarlaender/ [Accessed 22 September 2021].

Syrovátka, J. (2019). In Scrooge's boots: Lessons learned on disinformation from the 2019 European elections. *European View*, 18(2), 203–209.

The Associated Press. (2020). Report: Russia-linked disinformation operation still active. *The New York Times*. https://www.nytimes.com/aponline/2020/06/16/business/ap-us-russia-disinformation-report-1st-ld-writethru.html? [Accessed 22 September 2021].

*The Guardian*. (2015). Editorial: The Guardian view on Russian propaganda: The truth is out there. *The Guardian*. https://www.theguardian.com/commentisfree/2015/mar/02/guardian-view-russian-propaganda-truth-out-there [Accessed 22 September 2021].

Thomas, T. (2004). Russia's reflexive control theory and the military. *Journal of Slavic Military Studies*, 17(2), 237–256.

Thornton, R. (2015). The changing nature of modern warfare: Responding to Russian information warfare. *The RUSI Journal*, 160(4), 40–48.

TRT. (2020a, 23 March). Advised to stay home: But what about the homeless in times of coronavirus? *TRT World*. https://www.trtworld.com/perspectives/advised-to-stay-home-but-what-about-the-homeless-in-times-of-coronavirus-34804 [Accessed 22 September 2021].

TRT. (2020b, 19 May). Turquía Está a la Cabeza De Las Naciones Que Han Superado Este Proceso Con El Mínimo Daño Posible [Turkey Leads Among Nations That Have Overcome This Process With the Least Possible Damage]. https://www.trt.net.tr/espanol/vida-y-salud/2020/05/19/turquia-esta-a-la-cabeza-de-las-naciones-que-han-superado-este-proceso-con-el-minimo-dano-posible-1419911 [Accessed 22 September 2021].

TRT World. (2020c, 23 March). Coronavirus fear cripples refugees living in unsanitary camps. *TRT World*, 23 March 2020. *TRT World*. https://www.trtworld.com/magazine/coronavirus-fear-cripples-refugees-living-in-unsanitary-camps-34736 [Accessed 22 September 2021].

TRT. (2020d, 23 May). Turquía Y Rusia Inician Consultas Para Desarrollar Una Vacuna Contra El Covid19 [Turkey and Russia Begin Consultations to Develop Covid19 Vaccine]. https://www.trt.net.tr/espanol/turquia/2020/05/23/turquia-y-rusia-inician-consultas-para-desarrollar-una-vacuna-contra-el-covid-19-1422236 [Accessed 22 September 2021].

TRT. (2020e, 19 May). USA Testen Turkishbeam-Bestrahlungssystem Gegen Covid-19 [USA Testing Turkishbeam Irradiation System Against Covid-19]. https://www.trtdeutsch.com/news-turkei/usa-testen-turkishbeam-bestrahlungssystem-gegen-covid-19-2018413 [Accessed 22 September 2021].

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., … & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: a review of the scientific literature* (19 March 2018). https://hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf [Accessed 22 September 2021].

Twitter. (2019). *Information operations directed at Hong Kong*. Twitter Safety, 19 August 2019. https://blog.twitter.com/en_us/topics/company/2019/information_operations_directed_at_Hong_Kong.html [Accessed 22 September 2021].

US Department of State. (1987). Soviet influence activities: A report on active measures and propaganda, 1986–87.

Van Puyvelde, D. & Brantly, A. F. (2019). *Cybersecurity: Politics, Governance and Conflict in Cyberspace*. New York: John Wiley & Sons.

Wardle, C. (2017). Fake news. It's complicated. *First Draft*, 16, 1–11.

Weedon, J., Nuland, W., & Stamos, A. (2017). *Information operations and Facebook*. Facebook Newsroom. Facebook: https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf [Accessed 22 September 2021].

Welch, B. (2008). *State of Confusion: Political Manipulation and the Assault on the American Mind*. New York: Thomas Dunne Books, St. Martin's Press.

Wirtz, J. J. (2015). Cyber war and strategic culture: The Russian integration of cyber power into grand strategy. In: CCDCOE (ed.), *Cyber War in Perspective: Russian Aggression against Ukraine*. Tallinn: CCDCOE, pp.29–36.

Xinhua News Agency. (2020a, 20 March). Xinhua headlines: China returns solidarity with Europe in COVID-19 battle. *Xinhua News*. http://www.xinhuanet.com/english/2020-03/20/c_138898996.htm [Accessed 22 September 2021].
Xinhua News Agency. (2020b, 24 May). Un "Virus Politique" attaquant et dénigrant la Chine se propage aux Etats-Unis [A "Political Virus" attacking and denigrating China is spreading in the United States]. *Xinhua News*. http://french.xinhuanet.com/2020-05/24/c_139083968.htm [Accessed 22 September 2021].

# 7. AI and civic tech: Engaging citizens in decision-making processes but not without risks

## INTRODUCTION

> In particular, we need to examine the democratic implications of networked governance, how the structures and patterns of the new governance affect the balance of bureaucratic and democratic ethos, and how this balance affects, both positively and negatively, the citizenship and democratic deficits. Likewise, we need to examine the role citizens can play in networked government and collaborative governance.
> (Nabatchi, 2010, p.390)

Collaborative governance describes the institutional arrangements that aim "to empower, enlighten, and engage citizens in the process of self-government" (Sirianni, 2006, p.39). One of its key objectives is to ensure that a broad diversity of stakeholders take part in "collective decision-making process[es] that [are] formal, consensus-oriented, and deliberative and that aim to make or implement public policy or manage public programs or assets" (Ansell & Gash, 2007, p.544). Collective governance highlights indeed "the role of the public in a collaborative management process" (Cooper, Bryer, & Meek, 2008, p.221). In the last decades, "[n]ew forms of organization are appearing that give responsibility to the individuals based on their contributions and participation" (Sidjanksi, 2000, p.203).

Digital technology and citizen participation have become increasingly intertwined in the 21st century (Tolbert & McNeal, 2003). Digital technologies provide technical solutions to increase the number of stakeholders in a policy and decision-making process. Developments in Europe have shown that digital technologies can offer new avenues to re-enchant democracy and overcome some of its most pressing challenges. Pioneering examples of civic tech have burgeoned in both contexts. A series of ad hoc grassroots initiatives and hackathons has blossomed over the past decade into a larger civic tech community of for-profit and nonprofit organizations and investors. Such forms of citizen participation emerge to express populations' demands for greater equity, solidarity and to denounce the inaction of politicians toward global issues. Street

protests and activism abound, in cities around the world and on the internet and on social media platforms. As Dalton (2008) argues about the decline of conventional forms of citizen participation (e.g. elections) in liberal democracies, "the trends in political activity represent changes in the style of political action, and not just changes in the level of participation" (p.94). In other words, citizen participation today encompasses a larger range of actions (e.g. street protests, boycotts, and civic tech).

Governments increasingly use artificial intelligence (AI) in their efforts to foster citizen engagement and increase their participation in policy-making processes. This chapter focuses on civic tech, which is the "technology that is explicitly leveraged to increase and deepen democratic participation" (Gilman, 2017, p.745). This definition distinguishes particularly well between technology used to strengthen citizen participation and technologies used primarily to modernize operations and services (discussed in previous chapter). Civic tech can be developed and managed by different actors, such as technology start-ups, public administrations, and other political groups. In other words, not all civic tech are led by governments. They can be based on open access technology solutions (e.g. dicidim.org) or proprietary ones (the French start-up Cap Collectif managed the citizen consultation "grand débat national" in France in 2019) (Mabi, 2019). This high degree of diversity means that there is no universally accepted definition of civic tech, which creates a degree of vagueness in specifying the boundaries of these technologies and practices, and in particular what is public and what is private. However, all civic tech initiatives pursue one similar purpose: enhancing the participation of a broader range of stakeholders in public affairs. For Microsoft (2014), "Broadly defined, civic tech ranges from engagement between the city government and its population on social platforms, all the way to enterprise solutions that offer deep government IT problem-solving."

This chapter will explore civic tech and more precisely their use AI. Not all civic tech use of AI. But this technology can be very helpful in this context. For instance, it can help make sense out of thousands and thousands of comments (unstructured text) submitted by citizens in response to an online consultation. However, AI also raises numerous concerns in particular in the context of citizen participation. As discussed in previous chapters, by adopting AI in their relationship with citizens, governments introduce a fuzzy (i.e. conceptual challenges), variable (i.e. ongoing developments and applications), often opaque (i.e. black box phenomenon) agent in the citizen–government relation with various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). Moreover, the technology and the data collected can be in the hands of the private sector. In this context, the impacts of this technology on this relationship remain challenging to foresee and consequently difficult to prepare for. The notions of trust, transparency,

accountability, and equity may well be strongly challenged by the growing use of AI. This is particularly concerning when it comes to efforts to foster citizen engagement.

This chapter first explores the burgeoning field of civic tech and then focuses on the use of AI in this context. It aims to shed light on the asymmetry of power in the design and use of civic tech and AI-based tools and tactics used at different stages of the policy-making process. The choices made by designers and developers of civic tech shape who, when, and how citizens can participate in these platforms. These choices are within the digital infrastructure and not often not visible to the most affected stakeholders.

## CIVIC TECH AND PARTICIPATORY POLICY MAKING

In an effort to refocus the actions of governments and public institutions around the beneficiaries, and more specifically the citizens, new reflections are being conducted on the design of their services and technologies offered (Allio, 2014; Brown & Wyatt, 2010), on the strategies and tactics of citizen mobilization mediated by information and communication technlogies (ICTs) (Gatautis, 2010; Peixoto & Fox, 2016), and on the offer of new services allowing for new value creation (Denhardt and Denhardt, 2015; Osborne, 2017). It is in this context of redefining the role of the citizen in the policy-making process and more generally the relationship between government and citizens that civic tech has emerged.

### Collaborative Methods to Enhance Trust

Since the beginning of this century, the European Union (EU) has funded the development of more than 70 e-governance or e-democracy projects at the local, national and European level (Prieto-Martín, de Marcos, & Martínez, 2011). These efforts are in line with the EU's strategy to foster citizen participation and social innovations through digital technologies (European Commission, 2021). As mentioned earlier, a growing number of governments and public administrations are leveraging digital technologies to improve the efficiency of their operations and services. Faced with a certain low level of trust among citizens in national and European institutions (Eurobarometer, 2017) and in particular because of their lack of listening to their concerns (Pew Research Center, 2014, governments and European institutions are turning to new forms of governance and more inclusive and innovative co-decision processes to meet this growing demand for participation.

It is indeed well established that the public sector is still lagging behind in digitizing its operations and services and in adopting co-creation methods (Nunes, Galvão, & Cunha, 2014). In their global survey *Digital Government*

*Transformation: The Journey to Government's Digital Transformation*. Deloitte found that governments are at very different stages in their digital transformation journey, but the overwhelming majority are in the early or developmental stages of that transformation: nearly 70% of the agencies surveyed said they were lagging behind the private sector. In terms of motivation, cost and budget pressures and citizen demands are by far the top two drivers of digital transformation, accounting for 75% of responses (Eggers & Bellman, 2015).

Civic tech aims to contribute to strengthening democracy and more specifically democratic participation through innovative and collaborative governance (Moore & Hartley, 2008; Sørensen & Torfing, 2011; Ansell & Gash, 2007). The most common narrative associated with civic tech refers indeed to more responsive governance and a more meaningful engagement with other citizens and stakeholders (Mayur, Sotsky, Gourley, & Houghton, 2013). Moreover, it responds to the importance of lifelong civic engagement of learning experiences that cultivate citizens' perception that they can make change (i.e. political efficacy) and their belief in having responsibilities vis-à-vis the public good (i.e. civic identity). Civic tech can be seen as resting on three main pillars: transparency and accountability to hold governments accountable, citizen–government interaction, and digital tools to make citizens everyday live a little easier daily life (Dietrich, 2015).

This new approach to the institution–citizen relationship differs fundamentally from more traditional approaches to engaging citizens in policy-making processes, which too often limit their participation to the adoption stage. This co-creative approach therefore aims to open up other stages of the policy-making process to citizen participation so that they can contribute to proposing and implementing new solutions, perhaps also more adapted to their actual needs. Many non-governmental organizations (NGOs), citizens, and businesses are developing and managing digital tools to enhance the transparency and responsiveness of governments and improve the lives of their communities (Rumbul, 2016b).

This new approach aims at transforming the role of citizens into collaborative methods and refers to the transformation of citizens and other stakeholders from passive spectators to active contributors (Skaržauskienė & Mačiulienė, 2020). Hilgers and Ihl (2010) propose three dimensions of citizen collaboration: (i) citizen ideation and innovation, which allows public institutions to benefit from the knowledge and creative potential of citizens (e.g. open innovation platforms), (ii) collaborative administration, allowing for the mobilization of citizens with the aim of improving existing public administrative processes, and finally (iii) collaborative democracy, which includes the emergence of new modes of collaboration in order to improve citizen participation in political processes.

Three types of innovation have contributed to the development of civic tech: a more connected society (citizens and organizations) through ICTs (more online citizen interaction increases innovation capacity and leads to innovative solutions that are more responsive to people's needs), access to more data thanks to open data policies of public institutions (leading to increased visibility of issues and needs through cross-analysis of newly available data), and a great plurality of digital collaboration modes (various forms of collective intelligence and opinion aggregation) (Maciuliene, 2014).

## Typology of Civic Tech

Civic tech encompasses a wide range of technologies and activities aimed at improving the way people interact with government and each other (Knight Foundation & Rita Allen Foundation, 2017). Most research is identifying and mapping existing initiatives in the global north. Verhulst (2015) for instance identifies five overlapping component areas of civic technologies in the USA: (1) responsive and efficient city services, (2) open data portals and open government data, (3) engagement platforms for government entities, (4) community-focused organizing services and (5) geo-based services and open mapping data. Different mapping of civic tech exists at the national (e.g. in France and Switzerland[1]) and international level (e.g. civictech.guide[2] and participedia[3]). More recently, mapping efforts extended to the global south (Peixoto & Sifry, 2017). Social and Political Sciences, Computer-Supported Cooperative Work (CSCW) and Human–Computer Interaction (HCI) communities have examined civic data practices (Boehner & DiSalvo, 2016) and software development processes in civic projects (Knutas, Palacin, Maccani, & Helfert, 2019), as well as how digital technologies can support civic engagement (Asad & Le Dantec, 2015) and mobilize communities (Savage, Monroy-Hernandez, & Höllerer, 2016).

It is important to distinguish between initiatives initiated by public institutions (top-down) and those bottom-up initiated by other actors (e.g. citizen movements), since they might face different challenges (Knight Foundation & Rita Allen Foundation, 2017). For instance, e-participation processes initiated by government can be subject to institutional biases and built-in preconceptions about what users need (Rumbul, 2016a).

Top-down initiatives correspond to those participatory platforms either developed internally (e.g. by an IT department of a government) or externally (by companies and universities most often). They foster participation of citizens in some decision-making processes through digital technologies (i.e. e-participation and civic tech). They can also provide access to large datasets, and offer new approaches to service design (e.g. design thinking, co-production of services) (Skaržauskienė & Mačiulienė, 2020). They encompass a variety

of technologies (Linders, 2012), including artificial intelligence, to respond to a growing demand to digitize public action (de Feraudy, 2019). For instance, a number of local governments have developed an AI-powered social bot to optimize the online interaction with citizens and respond to the most common questions. This digital imperative cumulates with the participatory imperative already weighing on the construction, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017).

There is indeed a growing demand to digitize existing operations, administrative processes, and services. However, these efforts, otherwise known as e-government (Gilman, 2017), and discussed in previous chapters, are to be distinguished from civic technologies, whose main objective is not operational efficiency and effectiveness, but rather to foster participation.

Bottom-up initiatives are based on platforms developed outside the control of the state. As Badger (2012) and Suri (2013) point out, bottom-up initiatives are not necessarily designed to be disruptive of the political system and traditional processes of citizen participation like voting. They are primarily intended to complement existing processes and channels of communication previously monopolized by governmental and intergovernmental institutions. "Bottom-up" community civic tech typically includes some forms of tech activism, community-focused organizing services (Mačiulienė & Skaržauskienė, 2020) and leveraging open data – and sometimes open-source software – to address challenges that may be invisible to or neglected by government in a collaborative, problem-centered way (David, McNutt, & Justice, 2018). In addition, blockchain-based collective tools and intercultural communities, such as the Robin Hood Co-op,[4] enable new forms of financing and the protection of the commons and horizontal project and by doing so mobilize new forms of political subjects (Leander, 2021).

In their paper, Skaržauskienė and Mačiulienė (2020) categorized civic tech according to several dimensions: objective, target audience, and methods employed. The first dimension contains seven categories:

1. Improving government functions: these civic techs respond to the objective of digitizing public services, in order to increase the efficiency of public administration operations and services and improve public decision-making processes.
2. Improving the quality of life: these civic techs aim to improve the daily life of citizens, and include for example health services and education.
3. Solving societal problems: these civic techs aim to raise awareness and contribute to finding solutions to current societal challenges (e.g. gender gap).
4. Strengthening democracy: these civic techs offer tools to improve citizen engagement and voting, as well as various forms of free speech in society.

5. Community building: these civic techs offer tools and tactics to create and mobilize online networks and communities.
6. Sustainable future and environment: these civic techs propose new technological solutions (e.g. applications for mobility sharing or maximizing the circularity of digital devices) to contribute to environmental protection.
7. Transparency and accountability: these civic techs make government data available in an open, accessible and understandable way with the goal of making governance more transparent and accountable.

The civic techs presented in this chapter are not intended to be an exhaustive reflection of the great diversity of these initiatives in terms of both functionality and objectives. However, they have been selected for their representative character of the role of the citizen in co-creation and participation in the policy-making process. These civic techs come from previous studies and databases on citizen mobilization, social and digital innovations, e-participation and civic tech of course. This was also complemented by a consultation of websites of research centers specialized in civic tech (participedia.org, GovTech100, Microsoft Civic graph, digitalsocial.eu, Nominet Trust, Knight Foundation Research) and a web search combining a variety of keywords related to civic tech.

## Civic Tech Challenges

Although it is now well established that the use of ICTs has many benefits in society (Baack, 2015; McNutt et al., 2016), they are of course not the solution to all the challenges of liberal democracies. Digital technology can improve the dissemination of information (Weber, 2004), enhance understanding and facilitate the coordination of actions among citizens (Kreijveld, 2010). But it should be noted here that technology is not the only factor that increases interactions between citizens (Zappia, 2011). It is crucial not to adopt a determinist and solutionist approach to digital technologies. Civic tech is indeed no panacea and faces several challenges associated either with the platform itself or the users.

Top-down civic tech initiatives can be used as a form of veiled rhetoric or as a political marketing strategy for politicians. The digital instruments were often pursued as an objective in itself, symbolizing modernity more than a desire to really transform participation. Civic tech is also often approached with a certain fetishism of functionalities (e.g. the possibility to "like" contributions) without a clear a priori needs analysis (Albarède, de Feraudy, Marcou, & Saujot, 2018). They can also be vulnerable to institutional biases and rationale, and the resulting tools may be built with inherent assumptions concerning the users' needs (Skaržauskienė & Mačiulienė, 2020).

Many online governmental initiatives for citizen consultations promise to promote civic participation, but in practice, politicians use closed source code platforms, controlled and monitored by their managers (Santini & Carvalho, 2019). Moreover, government feedback on proposals made by citizens through civic tech is sometimes absent in these co-creation processes (Dahl & Soss, 2014; Sandfort & Quick, 2015).

One of the challenges associated with these initiatives is also the difficulty of quantifying and evaluating their impact on policy processes (Bruns & Swift, 2011). The debates and information exchanges that these platforms facilitate may not lead to any concrete results in terms of public policy and very rarely result in innovative solutions, consensus among stakeholders, or collective action (Cobo, 2012). Regarding co-creation aided by digital technologies, initiatives in the context of public services have either failed (Chadwick, 2011) or yielded limited results (Coleman, 2005; Peart & Diaz, 2007).

Furthermore, behind the alleged participatory processes, other power structures can be hidden (Pickard, 2008) and acting in an authoritarian manner and in the interest of small groups. Moreover, Skaržauskienė and Mačiulienė (2020) show that "civic tech are mostly oriented towards citizens' communities and governmental organizations" and other stakeholders are rarely involved in platforms' activities. What is more, "most of the initiatives focus only on the formation of a societal 'voice' and do not emphasize the feedback from government and the importance of co-creative synergy" (p.7).

This limitation adds to the fact that many citizens still lack access to the internet and have limited digital skills. Moreover, many citizens may lack the critical awareness regarding the type of technology used, the actors developing and managing the platform, the actors supporting the initiative, the transparency and accountability of data processing, and questions of cybersecurity and data privacy. Civic tech's digital infrastructures may indeed be opaque to the users. Some civic tech use AI-powered data processing techniques, which function as a black box, and hinder the participation's transparency and accountability. Data processing may be biased either due to the algorithm itself or the data sample. Additionally, the nature of data collected requires high security and privacy levels, which may be hampered by legacy infrastructure and cybersecurity vulnerabilities.

## AI TO AUGMENT HUMAN DATA PROCESSING CAPACITY IN CIVIC TECH

AI is used by some civic tech to enhance some functionalities. As discussed by Skaržauskienė and Mačiulienė (2020), most civic tech intend to give a voice to part of the population. In this context, the first use of AI is to augment the human capacity of processing large datasets. In other words, AI is useful when

confronted with a large number of comments: it replaces the need for a human being to read each single comment made by citizens. Instead, AI natural language processing capacity provides the initiators of the civic tech with a first analysis of clusters of opinions. As Daugherty and Wilson (2018) argue, the power of technology lies in its capability to complement and augment humans. Two examples are presented below. First, the online citizen consultations run by the EU Commission at different stages of the EU policy. Second, the "Grand Débat" and the "Vrai Débat" organized by the French government on the one hand and the Gilets Jaunes movement on the other hand.

**Citizen Consultation by the EU Commission**

The European Commission is the only European institution with the right to initiate the legislative process. The ordinary legislative procedure (previously called co-decision) allows the European Parliament (EP) on an equal footing with the Council of the European Union (thereafter the Council) to vote and make amendments to a policy proposal drawn by the European Commission (thereafter the Commission). The process repeats itself until the Council and the EP finally agree on a version of the text. This procedure aims to put at the same level citizen representations and government representations. Indeed, Members of the EP are elected through universal elections every five years. The Council is constituted of government representatives. The ordinary legislative procedure applies to a large range of topics, such as economic governance, immigration, energy, transport, the environment and consumer protection. In fact, the grand majority of European legislation is adopted through this procedure. This co-decision procedure was first introduced by the Maastricht Treaty on European Union in 1992, and then further developed and made more effective by the Amsterdam Treaty in 1999. Ten years later, the Lisbon Treaty renamed co-decision as ordinary legislative procedure, to reflect the fact that it became the primary legislative decision-making system of the EU.

Saurugger (2010) describes the "participatory turn" of EU institutions, where the necessity to foster the participation of citizens and civil society organizations was progressively acknowledged by EU representatives and became EU language and was included in a large number of policy papers and official communications from the Commission. This is well illustrated by the European Commission's White Paper on Governance (European Commission, 2001), which suggested to include citizens and civil society organizations in

the policy-making processes. Their involvement was seen as instrumental to the legitimacy of the EU policy process.

> The aim should be to create a transnational "space" where citizens from different countries can discuss what they perceive as being the important challenges for the Union. This should help policy makers to stay in touch with European public opinion, and could guide them in identifying European projects which mobilize public support. (European Commission, 2001, p.10)

The 2009 Lisbon Treaty reasserted the primacy of representative democracy principle (Title II, Art. 8A 1 TEU) but included elements of participatory democracy (Lindner et al., 2016) such as:

- The institutions shall, by appropriate means, give citizens and representative associations the opportunity to make known and publicly exchange their views in all areas of Union action.
- The institutions shall maintain an open, transparent and regular dialogue with representative associations and civil society.
- The European Commission shall carry out broad consultations with parties concerned in order to ensure that the Union's actions are coherent and transparent (Title II, Art. 8B TEU).
- In addition, the 2007 Lisbon Treaty initiated the European Citizens' Initiative (ECI), which allows citizens to suggest to the Commission legal action (Title II, Art. 8B 4 TEU).
- The Union offers citizens the opportunity to participate in public consultations and provide input throughout the policy cycle through a variety of mechanisms. A dedicated section of the EU Commission website called "Have your say"[5] offers the opportunity to all citizens to have a say in the policy-making process initiated by the Commission. It presents a number of policy initiatives in development for citizens to comment on. A search functionality allows citizens and businesses to look for open and closed consultation initiatives according to several criteria including topic (e.g. climate action), the stage in the policy-making process (e.g. in preparation), and the type of act (e.g. legislative proposal) and document (e.g. impact assessment report). Each initiative is described in a summary and the full document to download.

Another form of consultation at the EU level is the Conference of the Future of Europe, which consists of "citizen-led series of debates and discussions that will enable people from across Europe to share their ideas and help shape our common future" (European Commission, n.d.-a). A dedicated website allows citizens to share their views about the future of Europe.[6] Several topics are available (e.g. climate change and the environment) and for each topic

citizens can organize events, take part in events, and share ideas online (e.g. "The European Union could set up a program for returnable food packaging made from recyclable material.")[7] Each idea can then be commented on and endorsed. AI is used here to translate automatically content in different languages.

Prior to the Conference, the EU Commission already engaged a direct dialogue with citizens: town hall-style debates called "The Citizens' Dialogues" took place since 2012. The first one was organized on 27 September 2012 in the Spanish port city of Cadiz with then Vice-President Viviane Reding (European Commission, 2018). The White Paper on the Future of Europe in March 2017 launched a new step in the direct engagement of citizens in the design of the future of Europe (European Commission, 2017). In total, the process gave voice to hundreds of thousands of citizens who took part in some 1,600 citizens' dialogues in 583 locations throughout EU Member States and through the online consultation launched on 9 May 2018 (European Commission, 2018).

For the online consultation, twelve open and closed questions were proposed in all EU languages. Over 87,000 participants answered. An AI was used to analyze all answers and make sense of them:

> To identify the clusters of themes in each of the open-ended questions, Latent Dirichlet allocation was used as a topic-modelling technique. In this approach every contribution is allocated to one or several topics. Topics are identified based on associations of key words in the corpus of text. The allocation of verbatims to topics was reviewed by the research team on a sample of 100 verbatims per topic. Refinements to the allocation of keywords to topics were subsequently proposed. The resulting analysis allocates each response to several topics. (European Commission, 2019a, p.43)

This natural language processing capacity of AI is indeed useful when hundreds of thousands of comments in text format are produced by participants. To read each contribution, and then identify opinion clusters would be highly intensive in human labor. In this case, the AI used to overcome the limited data-processing capacity of human beings support the participation of more citizens. The EU Commission would have to limit the number of participants otherwise.

The EU Commission has expressed early its interest in data and digital technologies for policy making: "data technologies are amongst the valuable tools that policy makers have at hand for informing the policy process, from identifying issues, to designing their intervention and monitoring results" (European Commission, 2019b, para. 1). It commissioned a study conducted in 2015 to explore "the opportunities that innovative data-driven approaches offer for evidence-informed policy making, including the relevant data

sources and technologies" (Poel et al., 2015). Among their conclusions, they identified that most data-driven initiatives are "descriptive statistics and trend analysis, with some experiments using (advanced) sentiment mining, profiling, predictive analytics and other recent tools." What is more, these data-driven approaches are mainly employed for agenda setting and problem analysis (e.g. opinion mining online), enhance accountability and participation (e.g. civic tech platforms) but very few data-driven initiatives focus on policy evaluation and impact assessment (Poel et al., 2015).

The EU launched a forum for EU citizens to discuss EU policies and digital topics about the future of Europe called "Futurium" (European Commission, n.d.-b). The Futurium platform was originally developed to host and curate policy visions and ideas around the future of Europe. However, it has gradually evolved into a platform for experimenting with new ways of policy making based on scientific evidence and stakeholder participation. Its use of AI is here as well limited to processing a large quantity of text collected randomly on social media platforms. What Accordino (2013) refers to as Policy making 3.0:

> The platform hosts an online foresight toolkit to facilitate the joint creation of ideas to help design future policies. It leverages the potential of social networks, open data, semantic and knowledge mining technologies as well as participatory brainstorming techniques to engage stakeholders and harness their views and creativity to better inform policies that matter to them. (Accordino, 2013, p.321)

However, the use of AI also raises questions about the legitimacy of the outcome. In their survey about the use of AI by the EU Commission, Starke and Lünich (2020) showed that respondents perceive independent algorithmic decision-making (ADM) about the European Union (EU) budget to be illegitimate. EU policy makers should exercise caution when incorporating ADM systems in the political decision-making process. ADM systems for far-reaching decisions, such as budgeting, may only be used to assist or inform human decision-makers rather than replacing them. In their study called Data4Policy and commissioned by the European Commission, Rubinstein et al. (2016) highlighted that data collection and data analytics are not necessarily well understood by policy makers and other stakeholders, which presents a high risk for data-driven policy-making approaches. In their discussion about algorithmic decision-making (ADM), Starke and Lünich (2020) identify three types of challenges to the legitimacy of AI-informed policy-making processes:

> (a) On the input dimension, citizens may lack insight into or influence over the criteria or data that intelligent algorithms use to make decisions. This may undermine fundamental democratic values such as civic participation or representation. (b) On the throughput dimension, citizens may be unable to comprehend the complex and often inscrutable logic that underpins algorithmic predictions, recommendations, or

> decisions. The corresponding opacity of the decision-making process may violate the due process principle, for example, that citizens receive explanations for political decisions and have the opportunity to file complaints or even go to court. (3) On the output dimension, citizens may fundamentally doubt whether ADM systems actually contribute to better and/or more efficient policy. This may conflict with key democratic principles, such as non-discrimination. (p.e16-5)

Moreover, the many consultations conducted by the European institutions on the basis of the Treaties suffer from an "elitist bias," in so far as "they tend to be monopolized by a minority of actors who are very knowledgeable about European issues" (Costa, 2010, p.128), which may cause a phenomenon of self-selection of participants. Committees and lobby groups are the primary beneficiaries of these processes, which do not correspond to the spirit or primary motivation of these tools (which is to bring the citizen closer to decision-making). A second bias is related to rules and procedures, which can guide discussions and therefore their results (thus undermine the legitimacy based on efficiency) or even lead to a "mock discussion" (Gaudin, 2007, p.88).

## Civic Tech in France: Le Vrai et le Grand Débats

In 2019, while the political crisis resulting from the Yellow Vests movement seems to persist, the French President is launching the "Grand Débat National." This citizen consultation consists of public meetings but also the creation of an online platform called "Grand Débat."[8] To respond to this initiative, representatives of the "Gilets Jaunes" decided to launch their own debate platform called the "Vrai Débat."[9] Very quickly, and particularly in view of the success of these platforms and the importance of the comments collected, the question of the analysis and synthesis of the data collected (the text that citizens have published on each of these platforms as well as the results of the face-to-face debates) arose (Brugidou, Suignard, Escoffier, & Charaudeau, 2020). Both platforms called on the same French start-up (Cap Collectif)[10] to benefit from an artificial intelligence-based solution to process these vast datasets (Cointet & Parasie, 2018). AI was used to aggregate, prioritize and rank a very large number of comments and proposals.

However, these two civic tech platforms offer some notable differences. While both platforms offer the possibility to propose ideas, the "Grand Débat" is limited to four themes chosen by the government, while the "real debate" offers eight themes, to which one must add a "free expression" category, where all subjects can be addressed.

Moreover, the participation modalities are different. On the site of the "Grand Débat," the participants can only answer a precise questionnaire, in a rather dirigiste logic. Participants can answer closed or open questions,

which frame and close the expression. Questions stem proposals present in the program of Macron of 2017. The reason that led the government to opt for a platform stripped of these deliberative functions seems clear: not to make visible, through a hierarchy based on the number of votes and amendments, the most popular proposals (Courant, 2019).

The "Vrai Débat" allows to contribute in a freer way. Each proposal must however have a title, an explanation and specify the theoretical advantages of the proposal. It is also possible to add sources to contextualize the proposal (Rozières, 2019).

The differences do not end there. When the proposal is submitted, there is no possibility to debate on the "Grand Débat" site. It is only possible to consult the thousands of other contributions, sorting them by chronological order. On the contrary, on the platform of the "Vrai Débat," it is possible to consult other proposals and to vote for or against (or to have a mixed opinion), or to propose arguments for or against. On the "Vrai Débat" platform, proposals are aggregated, which avoids that 50 people have to write 50 times the same idea in isolation. In addition, the contributions with the most favorable votes appear first with graphs showing their scores, which is not the case on the "Grand Débat" platform. On the latter, the only criterion for prioritizing "contributions" is random or by date (Courant, 2019).

These functionalities are precisely those recommended by the National Commission for Public Debate for the platform launched by the government ("Grand Débat"). In its report, published on 14 January, it evoked in particular these kinds of functionalities: "Each contribution is visible to all with the possibility for everyone to support proposals already made and to comment on them" and "The platform serves to inform widely about the holding of the debate, its modalities and to involve the general public via an online vote on proposals expressed by users."[11]

Courant (2019) argues that the French government made this choice to keep control of the debate. Indeed, the production of such a large mass of data without any prioritization criteria make it almost impossible to process other than by artificial intelligence and algorithms or by a large number of people working full time. It is indeed detrimental, especially for the credibility of the debate, but also in terms of legitimacy. By keeping the data processing completely opaque (i.e. without any possibility for citizens and journalists to see transparently the most popular proposals), citizens can only see the AI-based data processing with suspicious eyes, which is reflected in a poll done in 2019: 58% of French people doubt that the proposals that will emanate from the debates will change the government's policy and 54% that they will be returned in all transparency and impartiality.[12] This is precisely a risk that Starke and Lünich (2020) among others highlighted: "On the throughput dimension, citizens may be unable to comprehend the complex and often

inscrutable logic that underpins algorithmic predictions, recommendations, or decisions" (p.e16-5).

## ADDITIONAL USES OF AI FOR CIVIC TECH

Civic tech encompasses a large variety of initiatives and actors. Some of these actors use AI to process the unstructured text produced by participants as discussed previously. However, it is not the only use of AI in this context. In their report entitled "Mediation and artificial intelligence: Notes on the future of international conflict resolution," DiploFoundation (Höne, 2021) provide an overview of how AI can be used in the context of mediation. They identify three broad potential applications and uses that are relevant for our discussion about civic tech: knowledge management, identification of specific needs, and enhancing diversity and inclusive participation.

### Knowledge Management

AI can help civic tech participants access relevant information so that they can make informed decisions about policy making. Traditional digital search methods are not very useful when data is either too abundant and unstructured. Citizens do not necessarily have the skills and the time to make thorough searches about complex topics. The solution is to use AI to perform a selection of the most relevant information on their behalf. This type of "smart searches" go beyond traditional keyword search to "make information and knowledge already accumulated more readily available and easier to search" (Höne, 2021, p.11).

AI can link data available in unstructured form from different sources, including legal databases, websites, social media, multimedia, scientific publications from think tanks and universities, among others. The ability to sort and aggregate the most relevant content represents a time and efficiency saving for civic tech participants who wish to give their opinion on a new policy for example. This type of advanced search would also reduce the information asymmetry between citizens and initiators. Faced with complex issues, and disinformation campaigns that have become commonplace on social media, this use of AI would be beneficial.

"Smart" or "intelligent" searches offer a number of additional benefits particularly relevant to civic tech users. First, it can understand human language. Since civic tech is used in many different contexts and fields, data can be written in domain-specific terminology. AI and in particularly NLP can increasingly understand the "linguistic nuances, synonyms, and relations found in everyday language and inside complex documents" (IBM, 2021) such as for instance legal and policy documents. Second, AI can develop an

understanding of document structures (e.g. elements such as headers, footers, charts, and tables), which differ from one source to another. This is particularly useful to highlight key arguments and extract content from documents from different fields and locations. Third, machine learning improves its capacity and precision with every new search, which can make this tool increasingly valuable. Last, it allows to organize the results according to a set of criteria relevant to the end user (IBM, 2021). These solutions are often used by large organizations to facilitate the access to information to their employees.

An "out of the shelf" solution increasingly used is the AI-powered news aggregator, which collects data from all over the web and posts it in one location. These solutions are not as elaborate as smart or intelligence search described previously as they focus on published web and social media content. There are pure news aggregators (e.g. News 360[13] or Feedly[14]), more specialized ones, such as poll aggregator (e.g. FiveThirtyEight[15]), top search results from multiple search engines (e.g. Dogpile[16]), or social media aggregators (e.g. Curator[17]). Moreover, such smart recommender systems and search algorithms could be selectable or configurable by citizens. "Unlike conventional recommender systems driven by a per-click business model, citizens' recommender systems are run by citizens themselves and serve the society as a whole" (Yang, Sun, Bozzon, Zhang, & Larson, 2017, p.388).

Global-Regulation[18] is a huge database enabled by AI-powered automatic translation, which allows users to find similarities between national legislations in the world. Goltz (2017) argues that this search engine could contribute to enhance democratic practices by enhancing the access to information. Fioriglio (2019) gives an insight about the Digital agora platform where users first have access to relevant information about a specific legislative announcement. Thanks to AI, citizens can then participate in policy and law making by providing well-informed and precise inputs rather than a general and vague opinion.

Bozdag and Van Den Hoven (2015) drew up a list of automated tools to avoid the phenomenon of filter bubbles described in the previous chapter. Munson, Lee, and Resnick (2013) created a browser extension called Balancer "that showed users feedback about the political lean of their weekly and all time reading behaviors" (p.419). This software supports users in accessing a plurality of sources of information, and step out of their "comfort zone" to consult only sources that comfort their existing political views. Social bots can also be used.

A social bot or chat bot is a "software system, which can interact or 'chat' with a human user in natural language such as English" (Shawar & Atwell, 2007, p.29). They are increasingly used by governments as a form of e-government service to answer the most common questions of citizens (Androutsopoulou, Karacapilidis, Loukis, & Charalabidis, 2019) via the

governmental website, social media platforms and on the phone. Tavanapour, Poser, and Bittner (2019) experimented with the use of chatbots "to improve, both, the documentation of citizens' contributions during on-site workshops and the idea generation in web-based e-participation by deploying an automated solution with a conversational agent" (p.2).

However, they offer other possibilities, including to mitigate political polarization. Hwang, Pearce, and Nanis asserted that "swarms of bots could be used to heal broken connections between infighting social groups and bridge existing social gaps. Social bots could be deployed to leverage peer effects to promote more civic engagement and participation in elections" (2012, p.40). Moreover, the study of Graham and Ackland (2017) demonstrated that social bots can be deployed to increase the visibility of opinions and the diversity of citizen views on social media platforms.

However, it also poses many challenges. First of all, the use of NLP has technical limitations, particularly in understanding context, processing extensive and varied vocabularies, handling different meanings, and understanding wordplay and ambiguity (Höne, 2018). Second, here too, the question of human agency in accessing information arises, and conversely, the question of transparency and accountability of AI in the selection of the information made available. In order to ensure the credibility and legitimacy of this type of research, AI must remain transparent and under human oversight as recommended by the EU. The independent High-Level Expert Group on Artificial Intelligence (HLEG) recommends indeed "that the decisions made by an AI system can be understood and traced by human beings" (HLEG, 2019, p.18).

**Identification of Specific Needs**

One of the most prominent claims of civic tech is to give a voice to citizens and facilitate interaction between a government and citizens. Debates and interactions on civic tech aim to contribute to reveal specific and tangible needs of a local population that may not be yet addressed by a government or local administration. In this context, AI can support this identification process and reveal new needs from discussions on social media platforms, which can then be discussed on civic tech. Liu, Tang, and Chen (2020) characterize artificial intelligence, combined with crowdsourcing intelligence as a possible new "hybrid intelligence." Enhancing efficiency in public policy making, civil society participation, also transparency and accountability: artificial intelligence has many promises to engage citizens and rule makers in shaping policy and society according to their needs provided that common values are shared.

Sentiment analysis allows to curate subjective information from social media platforms (Batrinca & Treleaven, 2015). Analysis of content published on social media allows to find relevant information that would otherwise be

difficult to access, such as identifying key people, monitoring major societal groups and social movements, discovering commonalities between different discussion streams, identifying questions to ask for citizen consultations and preventing the emergence of conflict situations.

Fernández-Martínez, Lopez-Sanchez, Aguilar, Rubio, and Nemegyei (2018) describe the prototype CoGovern developed to merge citizens' ideas and preferences about local politics is meant to support participatory decision-making, and to prevent policy makers from "cherry-picking." Indeed, after participatory processes, sometimes policy makers rearrange, hierarchize citizens' claims and political preferences without further explication of their choices and selection criteria to consider citizens' suggestions. CoGovern is then useful to prevent cherry-picking and to foster optimization and gathering of citizens' political suggestions.

However, here again, this use of AI presents many challenges. On the one hand, the increased surveillance is not without consequences for the privacy of citizens. The collected data could indeed be used for non-democratic purposes, or even to censor certain dissonant voices. On the other hand, since the data collection is done only on social media, it represents only a part of the population. Moreover, conversations on social media platforms are often sensationalist and not necessarily representative of the general opinion. AI can reproduce pre-existing biases, which are particularly numerous online. The AI-powered chatbot "Tay," developed by Microsoft to engage in casual conversations on Twitter, quickly adopted the sexist and racist positions it detected in Twitter users (Vincent, 2016). As Mijatović (2018) argues:

> Machines function on the basis of what humans tell them. If a system is fed with human biases (conscious or unconscious) the result will inevitably be biased. The lack of diversity and inclusion in the design of AI systems is therefore a key concern: instead of making our decisions more objective, they could reinforce discrimination and prejudices by giving them an appearance of objectivity. (Mijatović, 2018)

In that context, sentiment and network analysis on social media can be useful as well as misleading. On the one hand, only a part of the population is active on social media platforms. One must also highlight here the wide variety of social media platforms audiences: from TikTok, Snapchat, to LinkedIn and Twitter, the audience varies greatly in terms of age and interest. On the other hand, content shared on social media platforms may be much more extremist than what people would argue and decide in another context, for instance a civic tech initiative. Hence, this biased data collection needs to be completed by face-to-face or non-digital studies in order to ensure a better representativeness of the results.

**Enhancing Diversity and Inclusive Participation**

To make their debates and proposals legitimate, civic techs that want to represent the general public must be truly representative. In this context, AI can contribute to give a greater place to stakeholders who are traditionally little represented or heard, such as women or certain minorities. For instance, the UN Department of Political and Peacebuilding Affairs' (DPPA) Middle East Division (MED) is developing a new instrument to "evaluate the public's receptivity to an aspect of a peace agreement" thanks to the AI-powered analysis of discussions taking place online (i.e. digital focus groups) in various Arab dialects. Their objective is to allow "thousands of members of a concerned constituency in a country and its diaspora (e.g. refugees) to be consulted in real time" (UNDPPA, 2019). In the context of civic tech, one could envisage a similar instrument to ensure that a broader part of the population has a say when discussing a new policy or urban planning.

However, not all the population is connected to the internet, particularly elderly populations. Hence it is crucial to broaden the sources of data collection, and include other media such as the radio. For instance, UN Global Pulse used AI to analyze radio conversations on public policy and governmental initiatives in Uganda. Their objective was to hear from more than half of the population that uses radio as their primary source of information and to call in to share their views (Rosenthal, 2019). Thanks to text-to-speech AI applications to convert spoken words into text, UN Global Pulse could ensure a better representativity of the population. We could envisage the same approach for civic tech, where an AI can help gather in-person discussions for instance, or offer a phone number where people can record their opinion, which would be then converted into text and added to the rest of the data collected. An illustration is IBM's Project Debater Speech by Crowd, "an AI cloud technology for collecting free-text arguments from large audiences on debatable topics to generate meaningful narratives that express the participants' opinions in a concise way" (Ein-Dor, Gretz, & Bilu, 2019).

At the IBM THINK conference in Tel Aviv, which hosts more than 1,000 people, IBM researchers asked each attendee for their opinion on marijuana legalization. Then Speech by Crowd collected the opinions and created several narratives based on the pro and con opinions shared by the participants. It extracted the main arguments, summarized them concisely, and selected the highest quality arguments to support each topic. He then created concrete narratives to illustrate both sides of the debate, including the thinking of many different people. Then the researchers asked the audience again for their opinion on the same issue. The result showed that the majority of participants supported legalization at 76% (Curioni, 2019).

Another project also illustrates the promise of AI for better inclusion of different parts of the population in debates and in particular civic tech. The city of Lugano (Switzerland) called on IBM to better understand the position of citizens on the subject of public funding of autonomous vehicles. For 15 days, IBM and the city of Lugano collected citizens' opinions. They received more than 2,400 arguments. Speech by Crowd identified arguments for (including accident reduction, helping the elderly, and environmental protection) and against (the technology is immature, it will lead to a regression of the human condition, and the funding should instead be used to improve public transportation and job losses). Speech by Crowd also highlighted some of the arguments submitted by participants but omitted as outliers by the crowd, including reducing traffic. In the end, 68% of citizens voted in favor of funding the development of autonomous vehicles, and 32% voted against (Curioni, 2019). Similarly, the application Add-up is meant to facilitate public deliberation in the context of face-to-face debates. Plüss et al. (2018) propose this application with the purpose to create and catalyze agreement between users, "The goal of the ADD-up project is to transfer the advantages of deliberation support systems to face-to-face dialogic deliberations" (p.471).

In any civic innovation and civic tech, interaction design is crucial to enable the largest number of citizens to participate. It must be considered with great care. In particular, the choice of technology and the features that this technology allows must be carefully considered in order to avoid reinforcing existing discriminations, such as the digital divide. While research on Human–Technology Interaction (HCI) can adopt a technical or industrial perspective (Teli, Bordin, Blanco, Orabona, & De Angeli, 2015), recent developments focus in particular on the lived experience of people (Bødker, 2006). This research and the knowledge gained in this field can be useful to mobilize citizens in bottom-up creative practices (Schouten, Ferri, de Lange, & Millenaar, 2017; Cohendet, Grandadam, & Simon, 2010) and inclusive civic tech design and overcome barriers to citizen participation. Art is a motivating factor for urban transformation (Zukin, 1995). Mobilizing citizens' creativity can be an effective approach to reinvigorating civic engagement, especially when the sense of community is fading in rapidly growing cities (McAuliffe, 2012). Research has shown a significant correlation between culture, citizen participation and creativity (Varbanova, 2007).

A creative form of citizen engagement is pursued in the work of Li, Wang, Wang, Greuter, and Mueller (2020) who suggest using artificial intelligence to support citizen engagement through art. By promoting collaborative exchanges between citizens in the public space, street art (augmented by artificial intelligence) aims to reconnect not only citizens but also policy makers and their constituents with their urban environment. Through a joint civic and cultural engagement, Li, Wang, Wang, Greuter, and Mueller (2020) showed

that AI and street art could be promising avenues to foster civic engagement in policy making and urban planning. Other creative forms of citizen engagement include a playful participative installation that Sargeant, Dwyer, and Mueller (2018) have named "The Storytelling Machine." It transforms the public's drawings into animated characters moving through different digital worlds. With each new participant, the system builds a collective story. Moreover, Carter, Churchill, Denoue, Helfman, and Nelson (2004) built a digital graffiti system that enables participants to post digital graffiti annotations on a digital community bulletin board located in a public place. Similarly, Hoffman and Weinberg (2010) developed a robot powered by AI that can seamlessly adjust its improvisation and choreography when playing with a human musician concurrently.

The WeMonet tool aims to give creative, collaborative, and participative power to the citizen. First citizens add sketches to an online canvas, which "are synthesized, enhanced to be more vivid through machine learning algorithms, and projected on a screen, forming a participatory artwork" (Li, Wang, Wang, Greuter, & Mueller, 2020, p.1). This form of human–computer interaction promotes citizens' engagement in collaborative creative practices and enhances the city's creativity, and consequently makes the city more liveable and vibrant (Landry, 2012; Schacter, 2014). The concept of creative city explores how citizens think and act creatively, which can be viewed as a new approach to urban planning (Landry, 2012). It focuses also on citizens' lived experience instead of focusing primarily on infrastructure or the urban architecture (Landry & Bianchini, 1995; Varbanova, 2007). These studies are by far not representative of the breadth of variety of experiments in this field. But they show nevertheless that creativity can be an interesting avenue to explore to raise awareness about AI and develop new skills and literacy, as well as (re)mobilize citizens in projects related to their city or neighborhood.

## CONCLUDING REMARKS

As discussed in this chapter, civic tech refers to the technology that aims "to increase and deepen democratic participation" (Gilman, 2017, p.745). As Badger (2012) and Suri (2013) point out, they are primarily intended to complement conventional citizen participation and channels of communication previously monopolized by governmental and intergovernmental institutions, as well as address challenges that may be invisible to or neglected by government in a collaborative, problem-centered way (David, McNutt, & Justice, 2018). AI is used in this context for efficiency purposes: to process a vast number of comments and text published by citizens on some of these platforms. However, civic tech also presents challenges. First, many citizens still lack access to the internet and have limited digital skills, which means that "civic tools may

increase the divide and further marginalize those already limited in exerting power" (Skaržauskienė & Mačiulienė, 2020, p.11). Moreover, many citizens may lack the critical awareness regarding the type of technology used, the actors developing and managing the platform, the actors supporting the initiative, the transparency and accountability of data processing, and questions of cybersecurity and data privacy. Civic tech's digital infrastructures may indeed be opaque to the users. In addition, the growing role and influence of tech companies in the context of democratic processes and governance requires a close examination (Duberry, 2020).

When using AI, and because of its black box characteristics, it may be difficult to explain how AI makes its decisions. In other words, it could make the outcome document suspicious, that is, reducing trust in the process and its perceived legitimacy, as well as hinder citizen participation motivation. Data processing may also be biased either due to the algorithm itself or the data sample. In their study, Starke and Lünich (2020) showed that citizens could only perceive an AI-informed policy-making process as legitimate "when such systems operate under the scrutiny of democratically elected institutions (as in the hybrid condition)" (p.e16-13). Keeping human agency and oversight remains a precondition for implementing AI (Goldenfein, 2019), as also recommended in the EU strategy for trustworthy AI.

## NOTES

1. See the website of the Swiss Index of Civic Tech. https://www.epfl.ch/labs/lasur/fr/barometre-des-civic-tech-2019/ [Accessed 24 August 2021].
2. See the Civic Tech Guide website. https://civictech.guide [Accessed 24 August 2021].
3. See the Participedia website. https://participedia.net [Accessed 24 August 2021].
4. See the website of Robin Hood Co-op. https://www.robinhoodcoop.org [Accessed 24 August 2021].
5. See the EU Commission website "Have your say." https://ec.europa.eu/info/law/better-regulation/have-your-say_en [Accessed 24 August 2021].
6. See the website of Future EU. https://futureu.europa.eu/?locale=en [Accessed 24 August 2021].
7. Idea proposed by Francois Wekmans on 19 April 2021. https://futureu.europa.eu/processes/GreenDeal/f/1/proposals/83 [Accessed 24 August 2021].
8. See the platform of Grand Débat. https://granddebat.fr/ [Accessed 24 August 2021].
9. See the platform of Vrai Débat. https://le-vrai-debat.fr/ [Accessed 24 August 2021].
10. See the website of Cap Collectif. https://cap-collectif.com/ [Accessed 24 August 2021].
11. Translation of: "Chaque contribution est visible de tous avec la possibilité pour chacun de soutenir des propositions déjà émises et de les commenter." "La plateforme sert à informer largement de la tenue du débat, de ses modalités

et à faire participer le grand public via un vote en ligne sur des propositions exprimées par les usagers," from Joanno, C. (2019). Mission d'accompagnement et de Conseil pour Le Grand Débat National. Commission Nationale du Débat Public. https://www.archives.debatpublic.fr/sites/cndp.portail/files/documents/01-rapport-missiongd_ok-1.pdf [Accessed 24 August 2021].

12. Translation of "58% des Français doutent que les propositions qui émaneront des débats infléchiront la politique du gouvernement et 54% qu'elles seront restituées en toute transparence et impartialité," from Clavel, G. (2019). La popularité d'Emmanuel Macron penche toujours plus à droite [SONDAGE EXCLUSIF]. *Huffpost*. https://www.huffingtonpost.fr/amp/2019/02/06/la-popularite-demmanuel-macron-penche-toujours-plus-a-droite-sondage-exclusif_a_23663407/ [Accessed 24 August 2021].
13. See the website of News 360. https://news360.com [Accessed 24 August 2021].
14. See the website of Feedly. https://feedly.com [Accessed 24 August 2021].
15. See the website of FiveThirtyEight. https://fivethirtyeight.com [Accessed 24 August 2021].
16. See the website of Dogpile. https://www.dogpile.com [Accessed 24 August 2021].
17. See the website of Curator. https://curator.io [Accessed 24 August 2021].
18. See the website of Global-Regulation. https://www.global-regulation.com [Accessed 24 August 2021].

# REFERENCES

Accordino, F. (2013). The Futurium: A foresight platform for evidence-based and participatory policymaking. *Philosophy & Technology*, 26(3), 321–332.

Albarède, M., de Feraudy, T., Marcou, T., & Saujot, M. (2018). *Gouverner et innover dans la ville numérique réelle*. Audacities. IDDRI. https://fing.org/wp-content/uploads/2020/02/Audacities_Cas_CivicTechParticipation.pdf [Accessed 24 August 2021].

Allio, L. (2014). *Design Thinking for Public Service Excellence*. Singapore: UNDP Global Centre for Public Service Excellence.

Androutsopoulou, A., Karacapilidis, N., Loukis, E., & Charalabidis, Y. (2019). Transforming the communication between citizens and government through AI-guided chatbots. *Government Information Quarterly*, 36(2), 358–367.

Ansell, C. & Gash, A. (2007). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory*, 18(4), 543–571.

Asad, M. & Le Dantec, C. A. (2015). Illegitimate civic participation: Supporting community activists on the ground. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp.1694–1703).

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. *Big Data Society*, 2, 1–11.

Badger, E. (2012). The next big start-up wave: Civic technology. *CityLab*. http://www.citylab.com/tech/2012/06/next-big-start-wave-civic-technology/2265/ [Accessed 24 August 2021].

Batrinca, B. & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. *AI & Society*, 30(1), 89–116.

Bødker, S. (2006, October). When second wave HCI meets third wave challenges. In: *Proceedings of the 4th Nordic Conference on Human–Computer Interaction: Changing Roles* (pp.1–8).

Boehner, K. & DiSalvo, C. (2016). Data, design and civics: An exploratory study of civic tech. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp.2970–2981).

Bozdag, E. & Van Den Hoven, J. (2015). Breaking the filter bubble: Democracy and design. *Ethics and Information Technology*, 17(4), 249–265.

Brown, B. T. & Wyatt, J. (2010). Design thinking for social innovation. *Stanford Social Innovation Review Winter*, Stanford Graduate School of Business, pp.30–35. http://www.ssireview.org/images/articles/2010WI_Features_DesignThinking.pdf [accessed 24 August 2021].

Brugidou, M., Suignard, P., Escoffier, C., & Charaudeau, L. (2020). Un discours et un public "Gilets Jaunes" au coeur du Grand Débat National? Combinaison des approches IA et textométriques pour l'analyse de discours des plateformes "Grand Débat National" et "Vrai débat." In *JADT 2020: 15es Journées internationales d'Analyse statistique des Données Textuelles* (Vol. 1, pp.1–12).

Bruns, A. & Swift, A. (2011). g4c2c: A model for citizen engagement at arms' length from government. *JeDEM-eJournal of eDemocracy and Open Government*, 3(1), 57–69.

Carter, S., Churchill, E., Denoue, L., Helfman, J., & Nelson, L. (2004, April). Digital graffiti: Public annotation of multimedia content. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp.1207–1210).

Chadwick, A. (2011). Explaining the failure of an online citizen engagement initiative: The role of internal institutional variables. *Journal of Information Technology & Politics*, 8(1), 21–40.

Cobo, C. (2012). Networks for citizen consultation and citizen sourcing of expertise. *Contemporary Social Science*, 7(3), 283–304.

Cohendet, P., Grandadam, D., & Simon, L. (2010). The anatomy of the creative city. *Industry and Innovation*, 17(1), 91–111.

Cointet, J.-P. & Parasie, S. (2018). Ce que le big data fait à l'analyse sociologique des Textes, Un panorama critique des recherches contemporaines. *Revue française de sociologie*, 59(3), 533–557.

Coleman, S. (2005). The lonely citizen: Indirect representation in an age of networks. *Political Communication*, 22(2), 197–214.

Cooper, T. L., Bryer, T. A., & Meek, J. W. (2008). Outcomes achieved through citizen-centered collaborative public management. In: Blomgren Bingham, L. and O'Leary, R. (eds.), *Big Ideas in Collaborative Public Management*. Armonk, NY: M. E. Sharpe, pp.221–229.

Costa, O. (2010). The effectiveness at the cost of the order. The steady rationalization of the rules of procedure of the European Parliament. Centre d'études parlementaires; Université Luiss, July 2010, Rome, Italie. https://halshs.archives-ouvertes.fr/halshs-00541943 [Accessed 24 March 2022].

Courant, D. (2019). Petit bilan du Grand Débat national. *AOC média*, 1–7.

Curioni, A. (2019). IBM Project Debater demonstrates the future of democracy in Switzerland. *IBM Research blog*. https://www.ibm.com/blogs/research/2019/09/speech-by-crowd-switzerland/ [Accessed 24 August 2021].

Dahl, A. & Soss, J. (2014). Neoliberalism for the common good? Public value governance and the downsizing of democracy. *Public Administration Review*, 74(4), 496–504.

Dalton, R. J. (2008). Citizenship norms and the expansion of political participation. *Political Studies*, 56(1), 76–98.
Daugherty, P. R. & Wilson, H. J. (2018). *Human+ Machine: Reimagining Work in the Age of AI*. Brighton, MA: Harvard Business Press.
David, N., McNutt, J. G., & Justice, J. B. (2018). Smart cities, transparency, civic technology and reinventing government. In: Bolívar, M. P. R. (ed.), *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues.* Cham: Springer, pp.19–34.
de Feraudy, T. (2019). *Cartographie de la civic tech en France, Observatoire de la civic tech et de la démocratie numérique en France*, Décider ensemble. https://www.deciderensemble.com/page/300874-observatoire-civic-tech-et-democratie-numerique [Accessed 24 August 2021].
de Feraudy, T., & Saujot, M. (2017). Une ville plus contributive et durable: crowdsourcing urbain et participation citoyenne numérique. *Iddri Study*, 4, 1–72.
Denhardt, J. V. & Denhardt, R. B. (2015). *The New Public Service: Serving, Not Steering*. London: Routledge.
Dietrich, D. (2015). The role of civic tech communities in PSI reuse and open data policies. *European Public Sector Information Platform Topic Report*, 5.
Duberry, J. (2020). Le rôle grandissant des big tech dans la gouvernance environnementale. *The Conversation*. https://theconversation.com/le-role-grandissant-des-big-tech-dans-la-gouvernance-environnementale-150043 [Accessed 24 August 2021].
Eggers, W. & Bellman, J. (2015). *Digital Government Transformation: The Journey to Government's Digital Transformation*. Deloitte University Press. https://www2.deloitte.com/content/dam/insights/us/articles/digital-transformation-in-government/DUP_1081_Journey-to-govt-digital-future_MASTER.pdf [Accessed 24 August 2021].
Ein-Dor, L., Gretz, A., & Bilu, Y. (2020). IBM Project Debater. *IBM Research Blog*. https://www.ibm.com/blogs/research/2020/02/progressing-ibm-project-debater-at-aaai-20/ [Accessed 24 August 2021].
Eurobarometer. (2017). *Public Opinion in the European Union*. Brussels: European Commission.
European Commission. (2001). *European Governance. A White Paper*. COM 428 final, 25 July. Brussels: CEC.
European Commission. (2017). *White Paper on the Future of Europe: Reflections and Scenarios for the EU27 by 2025*. Publications Office of the European Union, Luxembourg. https://ec.europa.eu/info/sites/default/files/white_paper_on_the_future_of_europe_en.pdf [Accessed 24 August 2021].
European Commission. (2018). *Citizens' Dialogues and Citizens' Consultations. Key Conclusions*. Publications Office of the European Union, Luxembourg. https://ec.europa.eu/info/sites/default/files/euco-sibiu-citizensdialogues_en.pdf [Accessed 24 August 2021].
European Commission. (2019a). *Online Consultation on the Future of Europe Second Interim Report*. Publications Office of the European Union, Luxembourg. https://ec.europa.eu/info/sites/default/files/online-consultation-report-april-2019_en.pdf [Accessed 24 August 2021].
European Commission. (2019b). *Data4Policy*. https://www.data4policy.eu/ [Accessed 24 March 2022].
European Commission. (2021). Communication: "2030 Digital Compass: The European way for the Digital Decade." Communication from the Commission to the European parliament, the council, the European economic and social committee and the com-

mittee of the regions. https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52021DC0118&from=en [Accessed 22 March 2022].
European Commission. (n.d.-a). *Conference on the Future of Europe*. EU Commission website. https://ec.europa.eu/info/strategy/priorities-2019-2024/new-push-european-democracy/conference-future-europe_en [Accessed 24 August 2021].
European Commission. (n.d.-b) Futurium Platform. https://futurium.ec.europa.eu/en/discover-futurium/pages/about [Accessed 24 August 2021].
Fernández-Martínez, J. L., Lopez-Sanchez, M., Aguilar, J. A. R., Rubio, D. S., & Nemegyei, B. Z. (2018). Co-designing participatory tools for a new age: A proposal for combining collective and artificial intelligences. *International Journal of Public Administration in the Digital Age*, 5(4), 1–17.
Fioriglio, G. (2019). Automation, legislative production and modernization of the legislative machine: The new frontiers of artificial intelligence applied to law and e-democracy. In: Peruginelli, G. and Faro, S. (eds.), *Knowledge of the Law in the Big Data Age*. Sapienza Unviersità di Roma – Dipartimento di Scienze Politiche (Italy), p.60.
Gatautis, R. (2010). Creating public value through eParticipation: Wave project. *Economics & Management*, 15, 483–490.
Gaudin, J. P. (2007). *La démocratie participative*. Paris: Armand Colin.
Gilman, H. R. (2017). Civic tech for urban collaborative governance. *Political Science & Politics*, 50(3), 744–750.
Goldenfein, J. (2019). Algorithmic transparency and decision-making accountability. Thoughts for buying machine learning algorithms. In: Bertram, C., Gibson, A. and Nugent, A. (eds.), *Closer to the Machine: Technical, Social, and Legal Aspects of AI*. Melbourne: Office of the Victorian Information Commissioner, pp.41–60.
Goltz, N. (2017). Linked democracy 3.0-global machine translated legislation and compliance in the age of artificial intelligence. *SSRN 2977616*, p.1.
Graham, T. & Ackland, R. (2017). Do social bots dream of popping the filter bubble? The role of socialbots in promoting participatory democracy in social media. In: Bakardjieva, B. and Gehl, R. (eds.), *Socialbots and their Friend: Digital Media and the Automation of Sociality*. New York: Routledge, Chapter 10.
High-Level Expert Group on Artificial Intelligence (HLEG). (2019). *Ethics Guidelines for Trustworthy AI*. Publications Office of the European Union. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai [Accessed 24 August 2021].
Hilgers, D. & Ihl, C. (2010). Citizensourcing: Applying the concept of open innovation to the public sector. *International Journal of Public Participation*, 4(1), 68–88.
Hoffman, G. & Weinberg, G. (2010, May). Gesture-based human–robot jazz improvisation. In: *2010 IEEE international conference on robotics and automation* (pp.582–587). IEEE.
Höne, K. (2018). Cybermediation: What role for blockchain and artificial intelligence? *Diplo blog*, 12 October. https:// www.diplomacy.edu/blog/cybermediation-what-role-block-chain-and-artificial-intelligence [Accessed 24 August 2021].
Höne, K. (2021). Mediation and artificial intelligence: Notes on the future of international conflict resolution. *DiploFoundation*. https://www.diplomacy.edu/sites/default/files/Mediation_and_AI.pdf [Accessed 24 August 2021].
Hwang, T., Pearce, I., & Nanis, M. (2012). Socialbots: Voices from the fronts. *Interactions*, 19(2), 38–45.
IBM. (2021). Intelligent search. *IBM website*. https://www.ibm.com/cloud/learn/intelligent-search#toc-intelligen-oYpUPJxN [Accessed 24 August 2021].

Knight Foundation & Rita Allen Foundation. (2017). Scaling civic tech: Paths to a sustainable future. https://knightfoundation.org/wp-content/uploads/2020/03/Scaling_Civic_Tech_final.pdf [Accessed 24 August 2021].

Knutas, A., Palacin, V., Maccani, G., & Helfert, M. (2019). Software engineering in civic tech: A case study about code for Ireland. In: *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)* (pp.41–50). IEEE.

Kreijveld, M. (2010). *Unlocking the Full Potential of the Crowd: A Government Perspective*. London: Springer.

Landry, C. (2012). *The Creative City: A Toolkit for Urban Innovators*. London: Routledge.

Landry, C. & Bianchini, F. (1995). The creative city (Vol. 12). *Demos*, 13.

Leander, A. (2021). Redesigning the political with blockchain. In: Duberry et al., *Artificial Intelligence and Civil Society Participation in Policy-making Processes: Thinking about AI and Participation*. Proceedings to the workshop AI and civil society. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3817666 [Accessed 24 August 2021].

Li, Z., Wang, Y., Wang, W., Greuter, S., & Mueller, F. F. (2020). Empowering a creative city: Engage citizens in creating street art through human–AI collaboration. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, pp.1–8.

Linders, D. (2012). From e-government to we-government: Defining a typology for citizen coproduction in the age of social media. *Government Information Quarterly*, 29(4), 446–454.

Lindner, R., Daimer, S., Beckert, B., Heyen, N., Koehler, J., Teufel, B., … & Wydra, S. (2016). Addressing directionality: Orientation failure and the systems of innovation heuristic. Towards reflexive governance (No. 52). Fraunhofer ISI Discussion Papers. *Innovation Systems and Policy Analysis* n.52. https://www.isi.fraunhofer.de/content/dam/isi/dokumente/cci/innovation-systems-policy-analysis/2016/discussionpaper_52_2016.pdf [Accessed 22 March 2022].

Liu, H. K., Tang, M., & Chen, K. (2020). Public decision making: Connecting artificial intelligence and crowds. *The 21st Annual International Conference on Digital Government Research*, 15–19 June 2020, Seoul, Republic of Korea. New York : ACM, p.219.

Mabi, C. (2019). Grand débat : ce que la technique dit du politique. *Revue Projet*, (4), 20–24.

Maciuliene, M. (2014). Linking value co-creation and organizational absorptive capacity: Theoretical study and conceptual model. In: *Proceedings of the 14th International Academic Conference*, Valletta, Malta, 28–31 October 2014 (p.282).

Mayur, P., Sotsky, J., Gourley, S., & Houghton, D. (2013). *The Emergence of Civic Tech: Investments in a Growing Field*. Miami, FL: Knight Foundation.

McAuliffe, C. (2012). Graffiti or street art? Negotiating the moral geographies of the creative city. *Journal of Urban Affairs*, 34(2), 189–206.

McNutt, J. G., Justice, J. B., Melitski, J. M., Ahn, M. J., Siddiqui, S. R., Carter, D. T., & Kline, A. D. (2016). The diffusion of civic technology and open government in the United States. *Information Polity*, 21(2), 153–170.

Microsoft (2014). Civic tech: Solutions for governments and the communities they serve. *Microsoft Corporate Blogs*. http://blogs.microsoft.com/on-the-issues/2014/10/27/civic-tech-solutions-governments-communities-serve/ [accessed 24 August 2021].

Mijatović, D. (2018). In the era of artificial intelligence: Safeguarding human rights. *Open Democracy blog*, 3 July. https://www.opendemocracy.net/digitaliberties/dunja-mijatovi/in-era-of-artificial-intelligence-safeguarding-human-rights [Accessed 24 August 2021].

Moore, M. & Hartley, J. (2008). Innovations in governance. *Public Management Review*, 10(1), 3–20.

Munson, S., Lee, S., & Resnick, P. (2013). Encouraging reading of diverse political viewpoints with a browser widget. In: *Proceedings of The International AAAI Conference on Web and Social Media*.

Nabatchi, T. (2010). Addressing the citizenship and democratic deficits: Exploring the potential of deliberative democracy for public administration. *American Review of Public Administration*, 40(4), 376–399.

Nunes, A. A., Galvão, T., & e Cunha, J. F. (2014). Urban public transport service co-creation: Leveraging passenger's knowledge to enhance travel experience. *Procedia-Social and Behavioral Sciences*, 111, 577–585.

OECD. (2001). *Citizens as Partners: Information, Consultation and Public Participation in Policy-making*. Paris: OECD Publishing.

Osborne, S. P. (2017). Public management research over the decades: What are we writing about? *Public Management Review*, 19(2), 109–113.

Peart, M. N. & Diaz, J. R. (2007). Comparative project on local e-democracy initiatives in Europe and North America. *Research Centre on Direct Democracy, Faculty of Law, University of Geneva*. http://edc. Unige. Ch/download/ESF [Accessed 24 August 2021].

Peixoto, T. & Fox, J. (2016). When does ICT-enabled citizen voice lead to government responsiveness. In: Kaushik, B., Indermit, G., and Guislain, P. (eds.), *World Development Report 2016 Digital Dividends*. Washington, DC: World Bank Group.

Peixoto,T. & Sifry, M. S. (2017). *Civic Tech in the Global South: Assessing Technology for the Public Good*. Washington, DC: World Bank and Personal Democracy Press.

Pew Research Center. (2014). *A Fragile Rebound for EU Image on Eve of European Parliament Elections*. Washington, DC: Pew Global Publishing.

Pickard, V. W. (2008). Cooptation and cooperation: Institutional exemplars of democratic internet technology. *New Media & Society*, 10(4), 625–645.

Plüss, B., El-Assady, M., Sperrle, F., Gold, V., Budzynska, K., Hautli-Janisz, A., & Reed, C. (2018). ADD-up: Visual analytics for augmented deliberative democracy. *COMMA 2018*, *7th International Conference on Computational Models of Argument* (pp.471–472).

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., Mahieu, B., … & Svetachova, M. (2015). Data for policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking. *Report about the State-of-the-Art. Amsterdam: Technopolis*. Oxford Internet Institute, Center for European Policy Studies.

Prieto-Martín, P., de Marcos, L., & Martínez, J. J. (2011). The e-(R) evolution will not be funded. *European Journal of ePractice*, 15, 62–89.

Rosenthal, A. (2019). When old technology meets new: How UN Global Pulse is using radio and AI to leave no voice behind. *UN Global Pulse blog*. https://www.unglobalpulse.org/news/when-old-technology-meets-new-how-un-global-pulse-using-radio-and-ai-leave-no-voice-behind [Accessed 24 August 2021].

Rozières, G. (2019). Le "vrai débat" des gilets jaunes montre à quoi aurait dû ressembler le grand débat. *Huffpost*. https://www.huffingtonpost.fr/2019/01/30/

le-vrai-debat-des-gilets-jaunes-montre-a-quoi-aurait-du-ressembler-le-grand-debat_a_23656122/ [Accessed 24 August 2021].
Rubinstein, M., Meyer, E., Schroeder, R., Poel, M., Treperman, J., van Barneveld, J., … & Svetachova, M. (2016). Ten use cases of innovative data-driven approaches for policymaking at EU level. *Data & Policy*, 2, e16.
Rumbul, R. (2016a). Developing transparency through digital means? Examining institutional responses to civic technology in Latin America. *JeDEM-eJournal of eDemocracy and Open Government*, 8(3), 12–31.
Rumbul, R. (2016b). ICT and citizen efficacy: The role of civic technology in facilitating government accountability and citizen confidence. *IFIP World Information Technology Forum* (pp.213–222). Cham: Springer.
Sandfort, J. & Quick, K. S. (2015). Building deliberative capacity to create public value. In: Bryson, J. M., Crosby, B. C., and Bloomberg, L. (eds.), *Public Value and Public Administration.* Washington, DC: Georgetown University Press, pp.39–52. Santini, R. M. & Carvalho, H. (2019). The rise of participatory despotism: A systematic review of online platforms for political engagement. *Journal of Information, Communication and Ethics in Society*, 17(4), 422–437.
Sargeant, B., Dwyer, J., & Mueller, F. F. (2018, October). The Storytelling Machine: A playful participatory automated system featuring crowd-sourced story content. In: *Proceedings of the 2018 Annual Symposium on Computer–Human Interaction in Play Companion Extended Abstracts* (pp.285–294).
Saurugger, S. (2010). *Théories et concepts de l'intégration européenne*. Paris: Presses de sciences po.
Savage, S., Monroy-Hernandez, A., & Höllerer, T. (2016, February). Botivist: Calling volunteers to action using online bots. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp.813–822).
Schacter, R. (2014). The ugly truth: Street art, graffiti and the creative city. *Art & the Public Sphere*, 3(2), 161–176.
Schouten, B., Ferri, G., de Lange, M., & Millenaar, K. (2017). Games as strong concepts for city-making. In: Nijholt, A. (ed.), *Playable Cities*, Singapore: Springer, pp. 23–45.
Shawar, A. & Atwell, E. S. (2007). Chatbots: Are they really useful? *Journal for Language Technology and Computational Linguistics*, 22(1), 29–49.
Sidjanski, D. (2000). *The Federal Future of Europe: From the European Community to the European Union*. Ann Arbor, MI: University of Michigan Press.
Sirianni, C. (2006). *Reinvesting in Democracy*. Washington, DC: Brookings Institution Press.
Skaržauskienė, A. & Mačiulienė, M. (2020, December). Mapping international civic technologies platforms. *Informatics*, 7(4), 1–13.
Sørensen, E. & Torfing, J. (2011). Enhancing collaborative innovation in the public sector. *Administration and Society*, 43(8), 842–868.
Starke, C. & Lünich, M. (2020). Artificial intelligence for political decision-making in the European Union: Effects on citizens' perceptions of input, throughput, and output legitimacy. *Data & Policy*, 2.
Suri, M. (2013). From crowd-sourcing potholes to community policing: Applying interoperability theory to analyze the expansion of "Open311". *Berkman Center Research Publication* (2013-18).
Tavanapour, N., Poser, M., & Bittner, E. A. (2019). "Supporting the idea generation process in citizen participation – toward an interactive system with a conversational agent as facilitator." In: *Proceedings of the 27th European Conference on*

*Information Systems (ECIS)*, Stockholm & Uppsala, Sweden, 8–14 June 2019. ISBN 978-1-7336325-0-8 Research Papers. https://aisel.aisnet.org/ecis2019_rp/70 [Accessed 24 March 2022].

Teli, M., Bordin, S., Blanco, M. M., Orabona, G., & De Angeli, A. (2015). Public design of digital commons in urban places: A case study. *International Journal of Human–Computer Studies*, 81, 17–30.

Tolbert, C. J. & McNeal, R. S. (2003). Unraveling the effects of the Internet on political participation? *Political Research Quarterly*, 56(2), 175–185.

UN Department of Political and Peacebuilding Affairs and Centre for Humanitarian Dialogue (UNDPPA). (2019). *Digital Technologies and Mediation in Armed Conflict*. UN Press. https://peacemaker. un.org/sites/peacemaker.un.org/files/DigitalToolkitReport.pdf [Accessed 24 August 2021].

Varbanova, L. (2007). Our creative cities online. In: Svob-Dokic, N. (ed.), *Cultural Transition in Southeastern Europe: The Creative City – Crossing Visions and New Realities in the Region* (Culturelink Joint Publication Series, 11). Zagreb: Institute for International Relations, pp.9–18. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-58559 [Accessed 24 March 2022].

Verhulst, S. (2015). *Unpacking Civic Tech–Inside and Outside of Government*. New York: GovLab Digest.

Vincent, J. (2016). Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. *The Verge*. https://www.theverge.com/2016/3/24/11297050/tay-micro-soft-chatbot-racist [Accessed 24 August 2021].

Weber, S. (2004). *The Success of Open Source*. Cambridge, MA: Harvard University Press.

Yang, J., Sun, Z., Bozzon, A., Zhang, J., & Larson, M. (2017). CitRec 2017: International Workshop on Recommender Systems for Citizens. In: *Proceedings of the 11th ACM Conference on Recommender Systems, RecSys* 2017 (pp.388–389). Association for Computing Machinery (ACM).

Zappia, Z. (2011). *Participation, power, provenance: Mapping information flows in open data development* (Doctoral dissertation, Oxford Internet Institute).

Zukin, S. (1995). *The Cultures of Cities*. Cambridge, MA: Blackwell.

# Concluding remarks

This book explores AI-mediated citizen–government relations. It aims to answer the question: where and how is artificial intelligence (AI) used in this relation, which is essential for "the quality of democracy and strengthening civic capacity" (OECD, 2001, p.1).

Governments have progressively adopted a number of technology innovations to respond to a growing demand to (1) digitalize public action and optimize its operations and services (de Feraudy, 2019), and (2) increase citizen engagement in the development, implementation, and evaluation of public policies (de Feraudy & Saujot, 2017). The e-government efforts were mainly to take advantage of technological advances to (a) optimize the effectiveness and efficiency of government services, (b) put the citizen back at the center of the design of services rendered by organizations, and (c) increase trust in government (OECD, 2020). AI is increasingly used in the fields of healthcare, education, social and cultural services since it can be considered useful for six types of government challenges: allocating resources, analyzing large datasets, overcoming the shortage of experts, predicting scenarios, managing procedural and repetitive tasks, and diverse data aggregation and summarization (Mehr, Ash, & Fellow, 2017). The taxonomy developed by Misuraca and Van Noordt (2020) provides numerous examples of how AI is used by governments in Europe (European Union and United Kingdom, Norway and Switzerland) to improve knowledge management capacity (e.g. assist in the browsing and finding of relevant data in Slovakia), map and predict risks (e.g. predicts burglaries in Switzerland), and automatize data collection and analysis (e.g. process satellite imagery in Estonia), public services (e.g. self-driving snowploughs in Norway), decision-making processes (e.g. nursery child recruitment system used in Warsaw), and the communication with citizens (e.g. chatbot to answer frequently asked questions in Latvia).

But the citizen–government relation is more than the delivery of governmental services. It is also about including civil society in the policy-making cycle (OECD, 2001). The Multiple Streams Framework (MSF) is a powerful conceptualization of the policy process, and specifically agenda setting (Kingdon, [1984] 2011). It argues that policy entrepreneurs (e.g. civil society) need resources (e.g. technology) and specific skills (e.g. engaging multiple audience) to develop and implement tactics (e.g. narrative reframing) through problem, policy and politics streams, to identify and exploit successfully open

policy windows. Touraine (1992) contends that there cannot be any form of democracy without freedom of political choice. As Parry, Moyser, and Day (1992) contend, citizen participation corresponds to all these "action[s] by citizens which [are] aimed at influencing decisions which are, in most cases, ultimately taken by public representatives and officials" (p.16). If conventional forms of participation are in decline in some liberal democracies (Parvin, 2015, 2018), other forms of participation have developed including street protests and boycotts, leading some scholars to argue in favor of a transformation of citizen participation rather than a decline.

To strengthen citizen–government relations and citizen participation in policy making, the OECD (2001) recommends governments use digital technology for three types of actions: (1) enhancing access to information so that citizens are well informed, (2) enabling citizens to express their views on projects and societal issues that affect them in consultations, and (3) engaging citizens in decision-making processes. Information plays a crucial role throughout the policy-making process. Said differently, who provides and gains access to information, as well as who influences its distribution, gains a competitive advantage in the problem, policy and politics streams. Previous researchers have identified a number of key skills (Mintrom, 2019) and tactics (Goyal, Howlett, & Chindarkar, 2020) for policy entrepreneurs, many of which depend on access to information and information distribution capacity:

- collecting evidence to share (new and reliable) knowledge about alternatives, control information flows, and construct models of best practice;
- making arguments to alter problem perception, reframe a narrative or discourse, delegitimize institutional status quo, negotiate, and bargain;
- strategic thinking to exploit decision-making procedures;
- engaging multiple audiences to create awareness about a policy problem, politicize an issue, and mobilize public opinion.

Among the five policy spaces identified by Prateek, Kumar, Kar, and Krishnan (2021), online platforms constitute an unprecedented avenue to develop some of the tactics presented above and influence informally policy-making. These platforms offer civil society organizations and social movements an opportunity to develop creative advocacy campaigns to raise awareness, as well as to coordinate their actions (e.g. street protests). AI plays a key role on social media platforms in the form of their machine learning algorithms (MLAs), which controls information distribution with two primary objectives: keeping users online as long as possible, and overcoming the information overload. Although there is a large variety of platforms and MLAs, these two objectives are common to all. Based on the data collected from users, MLAs can predict with some degree of precision the information users will like the most

(Konstan & Riedl, 2012), and then rank, filter, and diffuse information accordingly. It leads to well-known phenomena such as filter bubble (Pariser, 2011) and echo chambers (Sunstein, 2001). In the case of Facebook, their MLA tends to favor juicy, sensationalist and extremist content, leading Schwartz (2015) to argue that this platform is not successful in creating a space for serious and critical interaction but rather echo chambers. AI enables online platforms to profile each user by processing big data collected from their online activity. Political communication has benefited from this opportunity to assess a potential voter's psychological characteristics (i.e. psychometric profiling) and micro-target them with individualized online ads. Beyond profiling and micro-targeting, this new generation of AI-powered tactics and tools includes programmatic advertising (AI placing ads online), political apps for smartphones, geotargeting services, automated profiles and social bots. Since AI also controls information diffusion, it also plays a key role in the diffusion of false news and in the mitigation of disinformation operations. The AI-powered tactics and tools used in the context of recent political elections in Europe, and examined in this book, are developed by a relatively small number of private sector companies, which benefit from the data and attention of large captive audiences. Their cost leads us to think that only stakeholders with substantial financial means can afford these tools, and the influence they offer. The limited transparency and accountability of these actors, and the tools they sell to some political actors, increase the asymmetry of power in the policy-making process, as well as raising questions about the legitimacy of the process itself. As Bradshaw and Howard (2019) argue, "Although there is nothing necessarily new about propaganda, the affordances of social networking technologies – algorithms, automation, and big data – change the scale, scope, and precision of how information is transmitted in the digital age" (p.11). AI and data are the two main components of this transformation.

These new tactics and tools are also used in the context of an "information warfare" (Thornton, 2015), and hybrid threat where information is the weapon and the minds of citizens the new "battlefield" (Cavelty & Mauer, 2016). Governments deploy cyber-capacity to weaken other states and intervene in their internal affairs through aggressive external cyber operations (Deibert & Pauly, 2019) in times of peace and conflict. Through disinformation campaigns abroad, some governments aim to influence sympathetic changes in citizen behavior and perception, erode trust and participation of some parts of the population in the decision-making process, decrease the quality of their communications environment, and diminish the quality of information availa-

ble to citizens (Krasodomski-Jones, Smith, Jones, Judson, & Miller, 2019). As Spruds et al. (2016) argue:

> [t]he factors that make this strategy so powerful are that this type of "warfare" is continuously ongoing and hard to detect. It is complicated to identify its source, particularly as more often than not it is waged from several sources simultaneously. And finally, such a warfare strategy penetrates all levels of society at a very low cost. Even if the audience does not necessarily believe in the planted information, the abundance of unvetted information of itself leads to a persistent distrust of public information and the media. (p.8)

AI is at the center of this battlefield both as an enabler of disinformation diffusion by controlling content distribution (i.e. MLA of online platforms favoring juicy content), and as a potential opportunity to mitigate their diffusion (i.e. automated fact-checking).

Disinformation operations are not the only threats to the trust between governments and citizens. Surveillance is an integral feature of online platforms (Trottier, 2016), where users watch over one another, states and intelligence agencies watch over a target population, and companies watch over their audience (Trottier, 2020). Surveillance and its impact on privacy and freedom of opinion and expression is well known. As mentioned previously, online platforms collect data from users' online activity, including demographics, psychographics, behavioral data, and metadata (i.e. data about data). Surveillance from governments can have two main purposes. On the one hand, intelligence agencies collect bulk data and use AI to identify potential threats to public safety and national security. On the other hand, governments and political parties use AI to analyze citizen conversations online (i.e. sentiment analysis and opinion mining) to understand what are the most pressing needs of their populations, as well as their opinions and arguments about specific topics (Milano, O'Sullivan, & Gavanelli, 2014). Lastly, surveillance is also conducted by private actors: as Zuboff (2015) argues, "[t]his new form of information capitalism aims to predict and modify human behavior as a means to produce revenue and market control" (p.75). The limited transparency and accountability of these forms of surveillance challenge the citizen–government relation. As Andrejevic (2007) argues, citizens have "limited knowledge about how their personal information is controlled, who controls it, and how it is used" (Andrejevic, 2007, p.27). This limited information and oversight contributes to an asymmetry of power in favor of those with surveillance means and to the erosion of trust between governments and citizens.

As mentioned previously, the OECD (2001) recommends using digital technology to offer civil society accessible and relevant consultation and decision opportunities. This new approach to the institution–citizen relations differs fundamentally from more traditional approaches to engaging citizens

in policy-making processes, which too often limit their participation to the adoption stage. This co-creative approach therefore opens up the problem and policy streams for civil society to play its role as policy entrepreneur. Civic tech refers to the technology that aims "to increase and deepen democratic participation" (Gilman, 2017, p.745). Top-down initiatives correspond to those participatory platforms either developed internally (e.g. by an IT department of a government) or externally (by companies and universities most often). But civic tech also comprises of bottom-up initiatives that are based on platforms developed outside the control of the state. As Badger (2012) and Suri (2013) point out, they are primarily intended to complement conventional citizen participation and channels of communication previously monopolized by governmental and intergovernmental institutions, as well as address challenges that may be invisible to or neglected by governments in a collaborative, problem-centered way (David, McNutt, & Justice, 2018). AI is used in this context for efficiency purposes: to process a vast number of comments and text published by citizens on some of these platforms. However, civic tech also presents challenges. First, many citizens still lack access to the internet and have limited digital skills, which means that "civic tools may increase the divide and further marginalize those already limited in exerting power" (Skaržauskienė and Mačiulienė, 2020, p.11). Moreover, many citizens may lack the critical awareness regarding the type of technology used, the tech actors developing and managing the platform, the political actors supporting the initiative, the transparency and accountability of data processing, and questions of cybersecurity and data privacy. Civic tech's digital infrastructures may indeed be opaque to the users. When using AI, and because of its black box characteristics, it may be difficult to explain how AI makes its decisions. In other words, it could make the outcome document suspicious, that is, reducing trust in the process and its perceived legitimacy, as well as hinder citizen participation motivation. Data processing may also be biased either due to the algorithm itself or the data sample. Additionally, the nature of data collected requires high security and privacy levels, which may be hampered by legacy infrastructure and cybersecurity vulnerabilities.

As discussed previously, AI is increasingly present in the citizen–government relation today. It mediates many interactions between civil society and those in power. The design justice perspective (Costanza-Chock, 2020) offers useful lenses to look at technology and the domination patterns it perpetuates. AI offers indeed a competitive advantage to those who have the financial means and/or the technological capacity to harness its power. In the politics stream, AI is mainly in the hands of governments, political leaders and parties, and interest groups with substantial financial means. Hence it reduces the spectrum of who can substantially influence policy making to these who can benefit from AI and big data competitive advantage. 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, 2007, 2003). This book concurs with these studies by arguing that in an AI-mediated citizen–government context, power lies with those who hold the AI and big data capacity, that is, not marginalized populations and individual citizens.

The introduction of AI in the citizen–government relations presents many opportunities but also many risks. AI remains indeed this (1) blurry (i.e. conceptual challenges, ongoing developments and multiple applications), (2) sometimes unreliable (i.e. AI technical or adversarial vulnerabilities, data and algorithm bias), (3) and often opaque (i.e. black box phenomenon) technological agent (4) with various degrees of agency (i.e. capacity to observe its environment, learn from it, and take smart action or propose decisions). When introducing (or allowing the use of) this technology in democratic processes, governments also introduce a degree of uncertainty and vulnerability. In other words, using AI in democratic processes is not neutral and may have long-lasting negative effects on the trust that citizens place in their governments, the transparency and accountability of policy making, as well as in their capacity to have a meaningful input in this process.

Civil society, and particularly individual citizens, take a risk when their interaction with government or their participation in policy making is mediated by AI. Tulloch and Lupton (2003) argue that voluntary risk-taking is this "activity in which individuals engage, and which is perceived by them to be in some sense risky, but is undertaken deliberately and from choice" (pp.10–11). This definition highlights three important elements: reflexivity (or consciousness) that one is taking a risk, capacity (or agency) to make the decision to take the risk, and voluntary aspect of the decision, which is shaped by social conditions to some extent (Zinn, 2015). In a context of constant technological, environmental and social transformations, it is crucial for citizens to develop their capacity to perceive and to respond to risk (Beck, 2009). However, as this book highlighted, the AI-mediation of citizen–government relations remains often blurry if not opaque to the citizen. This leads us to question the "voluntary" aspect of the risk-taking role of civil society. Hence, there is a dire need for capacity building (i.e. digital skills and literacy) to empower civil society and in particular individual citizens, so that they can adapt and benefit from this new AI-mediated citizen–government relations.

What is more, this book argues that governments may become risk-makers when introducing AI in their interactions with citizens, if this introduction is not done according to principles of equality, freedom, and human rights. The risk-taker role differs indeed from the risk-maker in the sense that the decision-maker is the one affected by the consequences of its decision (vs. affecting others). When adopting new technologies, and especially when the new technology is not mature in its development, early adopters may face mistakes, which then may jeopardize the confidence of later adopters in the technology (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). To avoid placing itself in the role of risk-maker (and consequently challenging the legitimacy of its role and decisions), governments must by all means ensure that (i) AI developers and managers apply principles of equality, freedom, and human rights, that (ii) citizen participation is not harmed by AI (i.e. discrimination against parts of the population, asymmetry of power in access to information and information distribution), and that (iii) the use of this technology also empowers civil society and greater inclusiveness in policy making and governmental services (i.e. not only efficiency). Otherwise, the introduction of AI may change how citizens perceive their agency and their role in the citizen–government relation.

As discussed previously, scholars have attempted to explore how and when a society becomes another (Koselleck, 1979; Castoriadis, 1997). Lefort (1988a) examined the transformational role of imaginary in politics and argued that a new political system emerges with the "mutation of the symbolic order" (Lefort, 1986, p.284).

Popular sovereignty structures the political imaginary of democracy (Diehl, 2019) and forms a "symbolic matrix of democracy" (Lefort, 1986). The principles of equality, freedom, and human rights are the criteria that legitimize political power, and become the normative horizon of democracy (Diehl, 2019).

In the case of AI, this technology presents many opportunities but also many risks. This is the paradox facing the governments of liberal democracies: supporting the development of a technology that will play a key role in the coming years (for society and economy to benefit from its promises), while ensuring that this technology does not contradict core values of liberal democracies. But the balance is fragile. And in dealing with a technology that presents so many risks for citizen participation, governments must avoid "manifesting a backlash concerning democratic values, [which] can indeed be the beginning of an erosion process and contribute to a mutation of the political imaginary" (Diehl, 2019, p.412).

The EU's approach to trustworthy AI goes in this direction:

> The Commission has developed key principles to guide the European approach to AI that take into account the social and environmental impact of AI technologies. They include a human-centric way of developing and using AI, the protection of EU values and fundamental rights such as non-discrimination, privacy and data protection, and the sustainable and efficient use of resources. (European Union, 2021)

This human-centric way of developing and using AI is indeed an imperative. But this book goes further. In light of the risks and challenges presented, a dedicated approach to AI for the citizen–government relation is needed, and in particular for citizen participation. In this context, each use of AI should not only adopt a human-centric approach, but also undergo a specific risk assessment to ensure the defense of equality, freedom, human rights, and the notion of popular sovereignty. Otherwise, it may lead citizens to conceive politics and governments differently, testifying of "a new system of representations" (Lefort, 1986, p.284). The risk is real. Without a distinct approach to AI for citizen participation, we might soon be mutating toward a different type of political system. But which one?

## REFERENCES

Andrejevic, M. (2007). *iSpy: Surveillance and Power in the Interactive Era*. Lawrence, KS: University Press of Kansas.

Badger, E. (2012). The next big start-up wave: Civic technology. *CityLab*. http: //www .citylab.com/tech/2012/06/next-big-start-wave-civic-technology/2265/ [Accessed 24 August 2021].

Beck, U. (2009). *World at Risk*. Cambridge: Polity Press.

Bradshaw, S. & Howard, P. N. (2019). *The global disinformation order: 2019 global inventory of organised social media manipulation*. Project on Computational Propaganda.

Castoriadis, C. (1997). The logic of the magma and the question of autonomy. In: Curtis, D. (ed.), *The Castoriadis Reader*. Oxford: Blackwell Publishing, pp.290–319.

Cavelty, M. D. & Mauer, V. (2016). *Power and Security in the Information Age: Investigating the Role of the State in Cyberspace*. London: Routledge.

Cinnamon, J. (2017). Social injustice in surveillance capitalism. *Surveillance & Society*, 15(5), 609–625.

Costanza-Chock, S. (2020). *Design Justice: Community-Led Practices to Build the Worlds we Need*. Cambridge, MA: The MIT Press.

David, N., McNutt, J. G., & Justice, J. B. (2018). Smart cities, transparency, civic technology and reinventing government. In: Bolívar, M. P. R. (ed.), *Smart Technologies for Smart Governments: Transparency, Efficiency and Organizational Issues.* Cham: Springer, pp.19–34.

de Feraudy, T. (2019). Cartographie de la civic tech en France, Observatoire de la civic tech et de la démocratie numérique en France, Décider ensemble.

de Feraudy, T. & Saujot, M. (2017). Une ville plus contributive et durable: crowdsourcing urbain et participation citoyenne numérique. *Iddri Study*, 4, 1–72.

Deibert, R. J. & Pauly, L. W. (2019). Mutual entanglement and complex sovereignty in cyberspace. In: Bigo, D., Isin, E., and Ruppert, E. (eds.), *Data Politics*. London: Routledge, pp.81–99.
Diehl, P. (2019). Temporality and the political imaginary in the dynamics of political representation. *Social Epistemology*, 33(5), 410–421.
Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human–Computer Studies*, 58(6), 697–718.
European Union (2021). *AI Excellence: Ensuring that AI works for people*. Digital Strategy. https://digital-strategy.ec.europa.eu/en/policies/ai-people [Accessed 30 September 2021].
Gilman, H. R. (2017). Civic tech for urban collaborative governance. *PS: Political Science & Politics*, 50(3), 744–750.
Goyal, N., Howlett, M., & Chindarkar, N. (2020). Who coupled which stream(s)? Policy entrepreneurship and innovation in the energy–water nexus in Gujarat, India. *Public Administration and Development*, 40(1), 49–64.
Harari, Y. N. (2018). *21 Lessons for the 21st Century*. London: Jonathan Cape.
Kingdon, J. W. ([1984] 2011). *Agendas, Alternatives, and Public Policies*. Boston: Little, Brown & Company.
Konstan, J. A. & Riedl, J. (2012). Recommender systems: From algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1–2), 101–123.
Koselleck, R. (1979). Einleitung. In: Brunner, O., Conze, W., and Koselleck, R. (eds.), *Geschichtliche Grundbegriffe* (Vol. 1). Stuttgart: Klett-Cotta, pp.viii–xxviii.
Krasodomski-Jones, A., Smith, J., Jones, E., Judson, E., & Miller, C. (2019). *Warring Songs: Information Operations in the Digital Age*. London: Demos Center.
Lefort, C. (1986). The logic of totalitarianism. In: Thompson, J. B. (ed.), *The Political Forms of Modern Society: Bureaucracy, Democracy, Totalitarianism*. Cambridge: Polity Press, pp.273–291.
Lefort, C. (1988a). The question of democracy. In: Lefort, C. (ed.), *Democracy and Political Theory*. Cambridge: Polity Press, pp.9–20.
Lefort, C. (1988b). "Interpreting revolution within the French Revolution. In: Lefort, C. (ed.), *Democracy and Political Theory*. Cambridge: Polity Press, pp.89–114.
Lyon, D. (2003). Surveillance as social sorting: Computer codes and mobile bodies. In: Lyon, D. (ed.), *Surveillance As Social Sorting: Privacy, Risk, and Digital Discrimination*. London: Routledge, pp.13–30.
Lyon, D. (2007). Surveillance, security and social sorting: Emerging research priorities. *International Criminal Justice Review*, 17(3), 161–170.
Mayer-Schönberger, V. & Ramge, T. (2018). *Reinventing Capitalism in the Age of Big Data*. London: Basic Books.
Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. *Ash Center for Democratic Governance and Innovation*, Harvard Kennedy School, August, 1–12.
Mintrom, M. (2019). So you want to be a policy entrepreneur? *Policy Design and Practice*, 2(4), 307–323.
Misuraca, G. & Van Noordt, C. (2020). AI Watch – artificial intelligence in public services: Overview of the use and impact of AI in public services in the EU. *JRC Working Papers* (JRC120399).
OECD. (2001). Engaging citizens in policy-making: Information, consultation and public participation. *Public Management Policy Brief*. Paris: OECD Publishing.

OECD. (2020). The OECD digital government policy framework: Six dimensions of a digital government. *OECD Public Governance Policy Papers*, No. 02. Paris: OECD Publishing.
Pariser, E. (2011). *The Filter Bubble: What the Internet is Hiding from You*. London: Penguin UK.
Parry, G., Moyser, G., & Day, N. (1992). *Political Participation and Democracy in Britain*. Cambridge: Cambridge University Press.
Parvin, P. (2015). Is deliberative democracy feasible? Political disengagement and trust in liberal democratic states. *The Monist*, 98(4), 407–423.
Parvin, P. (2018). Democracy without participation: A new politics for a disengaged era. *Res Publica*, 24(1), 31–52.
Prateek, G., Kumar, K., Kar, P., & Krishnan, A. (2021). Civil society as policy entrepreneur in agriculture and forestry sectors amidst COVID-19 lockdown in India. *Journal of Asian Public Policy*, 1–23.
Schwartz, S. A. (2015). Campaigning and contestation: Comments on politicians' Facebook pages during the 2011 Danish general election campaign. *Social Media+ Society*, 1(2), 1–11.
Skaržauskienė, A. & Mačiulienė, M. (2020, December). Mapping international civic technologies platforms. *Informatics*, 7(4), 46. Multidisciplinary Digital Publishing Institute.
Spruds, A., Rožukalne, A., Sedlenieks, K., Daugulis, M., Potjomkina, M., Tölgyesi, B., & Bruge, I. (2016). *Internet Trolling as a Hybrid Warfare Tool: The Case of Latvia. Results of the Study*. Latvian Institute of International Affairs Riga Stradins University.
Sunstein, C. R. (2001). *Republic.com*. Princeton, NJ: Princeton University Press.
Suri, M. (2013). From crowd-sourcing potholes to community policing: Applying interoperability theory to analyze the expansion of "Open311". *Berkman Center Research Publication* (2013-18).
Thornton, R. (2015). The changing nature of modern warfare: Responding to Russian information warfare. *The RUSI Journal*, 160(4), 40–48.
Touraine, A. (1992). What is democracy? *UNESCO Courier*. https://en.unesco.org/courier/novembre-1992/what-democracy [Accessed 21 August 2021].
Trottier, D. (2016). *Social Media as Surveillance: Rethinking Visibility in a Converging World*. London: Routledge.
Trottier, D. (2020). Denunciation and doxing: Towards a conceptual model of digital vigilantism. *Global Crime*, 21(3–4), 196–212.
Tulloch, J. & Lupton, D. (2003). *Risk and Everyday Life*. London: Sage.
Wu, T. (2010). *The Master Switch: The Rise and Fall of Information Empires*. New York: Vintage.
Zinn, J. O. (2015). Towards a better understanding of risk-taking: Key concepts, dimensions and perspectives. *Health, Risk & Society*, 17(2), 99–114.
Zuboff, S. (2015). Big other: Surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 30(1), 75–89.

# Index