

Faculté de droit et de criminologie

Artificial Intelligence-based Disinformation in Democratic Elections

International outlook and critical assessment of the
European Commission's 2021 draft Artificial
Intelligence Act

Auteure : Amélia CHOKOUALÉ DATOU

Promoteurs : Olivier DE SCHUTTER & Christophe LAZARO

Année académique 2020-2021

Master en droit à finalité droit de l'entreprise

page deliberately left blank.

Plagiat et erreur méthodologique grave

Le plagiat, fût-il de texte non soumis à droit d'auteur, entraîne l'application de la section 7 des articles 87 à 90 du règlement général des études et des examens.

Le plagiat consiste à utiliser des idées, un texte ou une œuvre, même partiellement, sans en mentionner précisément le nom de l'auteur et la source au moment et à l'endroit exact de chaque utilisation*.

En outre, la reproduction littérale de passages d'une œuvre sans les placer entre guillemets, quand bien même l'auteur et la source de cette œuvre seraient mentionnés, constitue une erreur méthodologique grave pouvant entraîner l'échec.

* A ce sujet, voy. notamment <http://www.uclouvain.be/plagiat>.

page deliberately left blank.

This paper has been written in English with the authorisation of Marie Aude Beernaert, President of the Master Panel, Charles-Hubert Born, Vice Dean and Christophe Lazaro, first Master thesis supervisor.

Table of contents

I.	Acknowledgements	9
1.	Introduction	11
2.	Elections within European democratic societies in crisis	13
2.1.	<i>From Athenian to modern democracy</i>	<i>13</i>
2.1.1.	Athenian direct democracy	13
2.1.2.	The 18 th century thinkers and representative democracy	14
2.1.3.	Modern democracy	15
2.2.	<i>Free and fair elections: the voice of the people</i>	<i>17</i>
2.3.	<i>The democratic crisis</i>	<i>18</i>
2.3.1.	The European integration	18
2.3.2.	The rise of extreme parties in Europe.....	19
3.	When the democratic debate occurs online.....	21
3.1.	<i>Social media and the democratic process</i>	<i>21</i>
3.1.1.	The revolution of Internet and social media.....	21
3.1.2.	Materializing the virtual debate	22
3.2.	<i>The social and algorithmic phenomena online</i>	<i>23</i>
3.2.1.	Capitalising on online engagement	23
3.2.2.	Human behaviour online	24
3.3.	<i>Intermediate conclusion</i>	<i>25</i>
4.	Manipulation of elections: disinformation and confusion	25
4.1.	<i>Disinformation and the right to free and fair elections.....</i>	<i>25</i>
4.2.	<i>Freedom of expression and the access to accurate information</i>	<i>26</i>
4.3.	<i>Disinformation disorders.....</i>	<i>27</i>
4.3.1.	The notion of disinformation.....	27
4.3.2.	Disinformation, not fake news	28
4.3.3.	The rationales of disinformation	29
	a. Disinformation through the years	29
	b. Disinformation disorder: symptom of other diseases	30

5.	Artificial intelligence: electoral manipulation through disinformation	31
5.1.	<i>Deepfake: a future tool or an overrated threat?</i>	<i>31</i>
5.1.1.	Deepfake in action	31
5.1.2.	The technology behind the fake	32
5.1.3.	Deepfakes and elections manipulation: the ongoing warning	33
	a. The skills required to create a deepfake.....	33
	b. Deepfake: real threat to elections?.....	34
5.1.4.	Deepfakes and fundamental rights	35
	a. The right to information and the Liar's Dividend	35
	b. The right to non-discrimination and freedom of press	36
5.2.	<i>Data collection and AI in political advertising</i>	<i>39</i>
5.2.1.	The Cambridge Analytica Scandal.....	39
5.2.2.	The impact of Cambridge Analytica	42
5.2.3.	A global issue	43
	a. Digital colonialism.....	43
	b. Brexit and the Leave.EU campaign.....	44
5.2.4.	A weapons-grade communication technique.....	44
5.2.5.	Targeted influenced campaign and fundamental rights.....	46
	a. Right to privacy	46
	b. Right to data protection and the principle of transparency.....	47
	c. Foreign intervention in national elections	49
5.3.	<i>Political bots.....</i>	<i>49</i>
5.3.1.	Political bots in action	49
5.3.2.	The technology behind the humanlike nature of bots	50
5.3.3.	Political bots and elections manipulation	51
	a. Large scale propaganda	52
	b. Altering the discussion equilibrium.....	52
	c. Generating a new narrative online	53
5.3.4.	The impact of political bots.....	53
5.3.5.	Political bots and fundamental rights	54
	a. Right to non-discrimination.....	54
	b. Fair and transparent electoral process	55
	c. Foreign intervention in national elections	55

5.4.	<i>Section conclusion: the growth of emotions-centred political strategy</i>	<i>56</i>
6.	Artificial Intelligence Act : perspectives for AI-related election manipulation.....	56
6.1.	<i>Context of the AI Act Proposal.....</i>	<i>57</i>
6.2.	<i>Objectives of the AI Act: the balance between innovation and safeguards.....</i>	<i>58</i>
6.3.	<i>The risk-based approach</i>	<i>59</i>
6.3.1.	Temporal, territorial and material scopes	59
6.3.2.	First harmonised definition of AI	60
6.3.1.	The pyramid of risks.....	60
	a. Unacceptable risks: prohibited AI practices	61
	b. High-risk and mandatory requirements	62
	c. Limited risks and the transparency requirement.....	63
6.4.	<i>AI election manipulation: Assessment of the AI Act draft.....</i>	<i>66</i>
6.4.1.	The challenge of articulation with other instruments	66
	a. Data protection and the compliance with the GDPR.....	66
	b. The Digital Services Act and illicit content online	68
6.4.2.	The lack of a societal risk against discrimination and electoral disruption.....	68
6.4.3.	Checks and balances of fundamental rights	69
6.4.4.	Foreign intervention in European and national democratic processes	70
6.4.5.	Transparency	70
7.	Conclusion	71
II.	Bibliography	74
III.	Glossary of terms	87

I. Acknowledgements

This Master thesis closes an important chapter of my academic years which I spent always well surrounded. For that I would like to express my appreciation to a few people who were there along the way: my first supervisor Christophe Lazaro who questioned my work in a sensitive, critical and relevant way; my second supervisor Olivier De Schutter who supportively took Dr. Lazaro's role when needed and, last but not least, my Module Leader of the European IT Law by Design course Rossana Ducato who inspired a number of reflections on AI.

I would like to acknowledge the input of my proof-reader, who also happens to be my father. On a similar note, I also would like to express my gratitude to my parents who endlessly care and encourage me in each of my endeavours, my sister Malaurie who is always available for a reassuring phone call or a remote sport session from Barcelona, my brother Lilian who helped me make enlightened decisions and who drew the pictograms of this paper based on my not-so-clear instructions. Together, we three siblings make a great team.

I also would like to thank my study buddies and friends Melanie and Pauline with whom I did countless hours on *Teams* during the last year and a half of pandemic. We created our new normal with these daily meetings.

Thank you to my roommates Shake, Danny, Nacho, Cam, BN, Wenzy, Pino and Swan for the social time away from my desk that I so much needed. Thank you to Solène, Alex, Princess, Soleil and Emeline. I wanted your names to be in here too.

Finally, I am also thankful to every individual with whom I had discussions about this project (however competent in politics, AI or technology). It is always in these settings that I found fresh perspectives on the subject.

The paradox of education is precisely this - that as one begins to become conscious one begins to examine the society in which he is being educated.

— **James Baldwin**

1. Introduction

“We need to understand the implication of deepfakes, the underlying AI technologies and the internet platforms that give them reach before considering the appropriate steps to mitigate the potential harms”, said A. B. Schiff during the 2019 United States House Intelligence Committee on deepfakes and artificial intelligence.¹ The 2020 U.S. elections were coming up and, following the experience of 2016, the fear of seeing a new election influenced by disinformation was rising. If disinformation in an electoral context has been extensively discussed in the U.S. in the wake of the Cambridge Analytica/Facebook scandal, the issue is a worldwide matter. In this regard, the European Union (hereinafter: “EU”) expressed concerns regarding false or misleading information spreading online and endangering democracy.² Guaranteeing the meaningful political participation of citizens entails safeguarding from any damaging interference the integrity of facts, fundamental rights, free and plural media coverage and an open democratic debate. Actors of elections disruption may be national or foreign, legal or natural persons, organisations or bodies. These actors now have access to disinformation weapons fit for the digital era. If Internet was first seen as an opportunity to connect individuals, opinions are now leaning more towards emphasizing its risks.³ Indeed, online-spread disinformation can derail the well-oiled democratic machine. Disinformation has existed for a long time, but Internet and social media boosted the possibilities. The political use of social media platforms such as Facebook or Twitter has shaped a new type of political campaigning where an emotional response is sought from citizens often to the detriment of them processing factually correct information.⁴ False information spreads easily online and it is even benefitting from the engagement-driven business model of social media.⁵ In fact, the social and algorithmic phenomena online tend to favour the engagement towards false and deceiving content. Today more than ever, an election disinformation campaign can be influential and result in behavioural change in the ballot.

¹ HOUSE INTELLIGENCE COMMITTEE, Member statement of Adam B. Schiff, National Security Challenges of Artificial Intelligence, Manipulated Media, and Deepfakes, 116th Congress, 2019, <<https://www.congress.gov/event/116th-congress/house-event/109620>> last accessed 3 June 2021.

² I. NENADIC, “Unpacking the « European approach » to tackling challenges of disinformation and political manipulation”, *Internet Policy Review*, 8(4), 2019, <<https://policyreview.info/pdf/policyreview-2019-4-1436.pdf>> last accessed 15 August 2021, p.3.

³ C. CADWALLADR, “Facebook’s role in Brexit and the threat to democracy”, Session 1, TED2019 Bigger Than Us, 15 April 2019, <https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy> last accessed 15 June 2021.

⁴ C. CADWALLADR, “Google, democracy and the truth about internet search”, *The Guardian*, 2016, <<https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook>> last accessed 30 June 2021.

⁵ N.I. BROWN, “Deepfakes and the Weaponization of Disinformation”, *Virginia Journal of Law & Technology*, 23(1), 2020, <https://static1.squarespace.com/static/5e793709295d7b60295b2d29/t/5f0b7e4e0f4f095f7e70613f/1594588750542/v23i1_1-Brown.pdf> last accessed 31 July 2021, p.19.

But what A. B. Schiff referred to in his opening statement before the U.S. House Intelligence Committee is not just any type of disinformation. The Committee was in fact discussing disinformation “*on steroids*”, meaning disinformation relying on artificial intelligence (hereinafter: “AI”).⁶ AI development is not just a question of innovation, it also establishes new dangers to social values and fundamental rights.⁷ AI can automate human decision-making behaviour. When discussing AI and elections, K. Manheim and L. Kaplan refer to AI as a democratic threat that can “*hack the vote*” and “*hack the mind*”.⁸ AI election manipulation tools combine AI technology, misinformation techniques and the broadcast power of social media.

This paper will analyse three AI tools of elections manipulation. First, deepfakes which are digitally altered videos of someone’s face or body. Second, targeted political propaganda which is a type of computational propaganda promoting political views on social media by injecting personalised content into the virtual bloodstream. And third, political bots which form another type of computational propaganda by disseminating disinformation through humanlike robots. The international community has just started to fight back against “fake news”, but manipulation tools powered by AI are already here. The objective of this work is to understand what we, as a society, are up against and to unfold one pending EU instrument to examine how this threat is taken into account in the draft policy.

This work is the result of a qualitative textual research and of dissecting the European Commission’s proposal for an artificial intelligence regulation. The objective is to conduct an interdisciplinary analysis of AI-driven disinformation tools by including legal, social and technological considerations. This paper often refers to U.S. examples and studies. The American influence is justified by the recent strong and illustrative examples of elections manipulation and the richness of the academic sources. However, this paper remains EU-centred and does not intent to conduct a comparative research. The structure is deliberately written in a “funnel-type”, meaning that it starts as a very broad problems statement and then proceed to analyse only the answer given by the Artificial Intelligence Act draft. The intention was to write a paper mirroring my vision of law: legislation drafting is always an attempt to solve issues human beings are facing as a community. This paper will not discuss AI as a solution against disinformation, nor discuss ethical design of AI

⁶ R. CHESNEY & D. CITRON, “Disinformation on Steroids: The Threat of Deep Fakes”, Council on Foreign Relations, 2018, <<https://www.cfr.org/report/deep-fake-disinformation-steroids>> last accessed 3 March 2021.

⁷ K. MANHEIM & L. KAPLAN, “Artificial Intelligence to Privacy and Democracy”, *The Yale Journal of Law and Technology*, 21, <https://yjolt.org/sites/default/files/21_yale_j.l._tech._106_0.pdf> last accessed 2 August 2021, p.138.

⁸ K. MANHEIM & L. KAPLAN, *op. cit.* (voy. note 7), p.138.

or present extensively the EU framework against disinformation. What this paper will discuss however is the historical evolution and the importance of elections within European democratic societies (section 2), the newly acquired role of Internet and social media in the democratic debate (section 3) and the functioning of disinformation and election manipulation (section 4). Thereafter, it will proceed to analyse three AI disinformation tools (deepfakes, targeted political campaigning and political bots) as well as their fundamental rights implications (section 5), and it will assess the 2021 Artificial Intelligence Act proposal of the Commission in light of these AI disinformation tools (section 6). This paper will ultimately answer the question: How is the European Artificial Intelligence Regulation proposal protecting free and fair democratic process from AI election manipulation?

2. Elections within European democratic societies in crisis

2.1. From Athenian to modern democracy

2.1.1. Athenian direct democracy

Originated in Athens in the fifth century BC, the word ‘democracy’ means ‘ruled by the people’. It comes from the terms: *demos* (the people) and *kratos* (the power).⁹ Democracy links two essential constituents of a state: the people and the government. And it entails that at a certain stage of the decision-making process a decision must come from a group of equal participants, the citizens. The Athenian democracy was centred around three revolutionary elements of direct democracy. First, to have Athenians directly involved into the political decisions, any citizens (namely any free men of twenty years old and over) could participate and vote at the Citizens Assembly (*Ecclesia*).¹⁰ Secondly, elections were organised to designate the *Magistracies* which were a sub-group in charge of the military and the public affairs of the city. Finally, citizens seating at the Council presiding the Assembly (*Boulè*) and at the Tribunal (*Heliaia*) were randomly picked. The Athenian conception of ruling at that time was pioneering and was rediscovered later when Europe needed an institutional change.

Athenian democracy also introduced the principle of equality. The group making a political decision was a group of equal citizens. However, if equality was at the core of Athenian democracy, in practice

⁹ C. MOSSE, *Regard sur la démocratie*, Paris, Perrin, 2013, pp.7-12.

¹⁰ F. MILLER, “Aristotle's Political Theory”, in *The Stanford Encyclopedia of Philosophy* (ed. E. N. ZALTA), Winter 2017 Edition, <<https://plato.stanford.edu/archives/win2017/entries/aristotle-politics/>> last accessed 17 November 2020.

the system was highly unequal. Women, foreigners and slaves were not considered citizens and were thus excluded from any deliberation. Moreover, mainly rich citizens living in the city could in practice spend time governing and discussing matters of the city at the *Ecclesia*. The three revolutionary elements of direct democracy were also challenged over time. The elected *Magistracies* became progressively the actual leaders of Athens, leaving the citizens with a modest part of the decision-making process. Among Ancient times authors and philosophers, the question of election manipulation already arose. Plato (428-348 BC) deemed democracy as fated to fail.¹¹ He justified his rather pessimistic assessment with two explanations. First, the principle of equality between participants exposes the decision-makers to the changing desiderata of the majority.¹² Second, the elected person will not necessarily be the one with the most expertise but rather the person most appealing to the population.¹³ Plato feared experts in “winning of elections” and “manipulation” will ultimately be governing the state. Aristotle (384-322 BC) was a strong advocate for political participation. According to Aristotle, a system is oligarchic if the people in power are elected, and democratic if they are randomly picked.¹⁴ However, he does consider that elections can tend towards more democracy if the entire population participates and votes. He wrote: “*Political justice exists between free and equal individuals who share in common a life which aims at self-sufficiency under the rule of law*”.¹⁵ Both Plato and Aristotle already pinpointed some of the dangers democracy might face: elections manipulation, demagoguery and the risk of electoral apathy.

2.1.2. The 18th century thinkers and representative democracy

In the 18th century, the idea of a state as a “*right-protecting and right-honouring institution*” arose.¹⁶ Societies who were willing to consider a democratic regime again had to adapt the ancient model. citizens assemblies were manageable within the limited size of Athens in the 5th century. But the Athenian direct democracy could not be replicated in the 18th century. Authors like Thomas Paine, Condorcet or Sieyès started to refer to the concept of “*representative democracy*” or “*representative*

¹¹ S. LANGE, “Plato and Democracy”, *The Classical Journal*, 34(8), 1939, <https://www.jstor.org/stable/3291135?seq=1#metadata_info_tab_contents>, pp.480-486.

¹² G. HERMET, « Les choix fondateurs », in *La Démocratie* (ed. G. HERMET), Paris, Flammarion, 1997, pp.14-16.

¹³ T. CHRISTIANO, « Democracy », in *The Stanford Encyclopedia of Philosophy* (ed. E.N. ZALTA), Fall 2018 Edition, <<https://plato.stanford.edu/archives/fall2018/entries/democracy/>>.

¹⁴ Y. SINTOMER, « Le tirage au sort à travers l’histoire : une domestication du hasard ? », in *Petite histoire de l’expérimentation démocratique. Tirage au sort et politique d’Athènes à nos jours* (ed. Y. SINTOMER), Paris, La Découverte, Poche/Essais, 2011, <<https://www-cairn-info.proxy.bib.ucl.ac.be:2443/petite-histoire-de-l-experimentation-democratique--9782707170149-page-39.htm>> p. 39-102.

¹⁵ T.C. LOCKWOOD JR., “Polity, Political Justice and Political Mixing”, *History of Political Thought*, 27(2), 2006, pp.207-222, <https://www.jstor.org/stable/26222194?seq=6#metadata_info_tab_contents> last accessed 5 June 2021, p. 209.

¹⁶ N. WOLTERSTORFF, *Understanding liberal democracy: Essays in political philosophy*, Oxford University Press, 2012, pp.2-6.

government”.¹⁷ Representative democracy enables citizens to elect representatives. They will be vested with the responsibility to rule for the people during a given period of time. This responsibility demands to discuss public matters, enact laws and decide policies. As in the case of the Athenian direct democracy, authors raised concerns regarding the system. According to Jean-Jacques Rousseau (1712-1778), there is at the foundation of sovereignty a consensus (*Social Contract*) existing between the rulers and the governed.¹⁸ The idea is that humankind shall cooperate to protect one another within a society while still maintaining individuals’ free will. However, he considered representative democracy as a dangerous regime in which the will of the people can be corrupted.¹⁹ In *Du Contrat Social*, Rousseau wrote that there was an inherent risk that representatives will put their own personal interests over the collective interest.

2.1.3. Modern democracy

There are vestiges of direct democracy in our modern representative democracy. The most common direct democracy method remains the referendum.²⁰ According to political scientist N. Urbinati, democracy has not so much evolved in theory since the institutional changes of the 18th century.²¹ But modern European democracy brought the universal suffrage. Democracy has always been defined as encompassing the principle of equality. In theory, every citizen should have the power to voice their opinions. However, through the years, democracy co-existed with non-equal outsets.²² Certain parts of the population including slave and women were thus excluded from the democratic process.

If in theory democracy has not seen an important evolution, in practice external factors influencing the functioning of societies at large have an impact. In the midst of the digital transformation, democracy and its principles are shaken by a risk that was already pinpointed by authors and philosophers throughout history: the risk of elections manipulation. Plato or Rousseau both already

¹⁷ H.LANDEMORE, *La démocratie représentative est-elle réellement démocratique*, La vie des idées, Interview of N.URBINATI & B. MANIN by H. LANDEMORE, <https://booksandideas.net/IMG/pdf/20080307_landemore.pdf> last accessed 2 June 2021, p. 2.

¹⁸ F. TINLAND, « Hobbes, Spinoza, Rousseau et la Formation de l’idée de démocratie comme mesure de la légitimité du pouvoir politique », *Revue Philosophique de la France et de l’Etranger*, 175(2), 1988, pp.195-222, <https://www.jstor.org/stable/pdf/41093850.pdf?casa_token=SpMAL5-iTF0AAAAA:t43x9tRPOfmZ7rSULKRLCV38z33Sd0EczAkyPXGu_vKwHbqJjMwZR5HfPCFaA-IN59aYg3FQBjYgiTfSXxOPW5WXnmNjOgji22LKAxHUSq765sogo4T6> last accessed 27 May 2021, p.197.

¹⁹ F. TINLAND, *op. cit.* (voy. note 18), p.200.

²⁰ J. G. MATSUSAKA, *Let the People Rule: How Direct Democracy Can Meet the Populist Challenge*, Princeton University Press, 2020, p. 5.

²¹ N.URBINATI, “Representative democracy and its critics”, in *The Future of Representative Democracy* (eds. S. ALONSO, J.KEANE, W. MERKEL), Cambridge University Press, 2011, pp. 23-49.

²² J. STAAL, *Propaganda Art in the 21st Century*, MIT Press, Cambridge, 2019, p.19.

feared the occurrence of fraudulent manoeuvres. However, thoughts on how to deal with people's opinions are divergent. French sociologist A. Touraine considers democracy as a matter of free choice of government and “*not the pursuit of “popular” policies*”.²³ Whereas researcher and democracy expert M. Rashbrooke considers that modern democracy should tend towards the increase of “everyday democracy”, appealing to the people.²⁴ The problem of election manipulation has been emphasized on many occasions throughout History. But as a reoccurring issue in democratic regimes, what are the effective solutions? The question is not on the type of regime in place. This paper is not questioning democracy but rather its existing safeguards. When discussing democracy, the United Nations (hereinafter: “UN”) advocates for sets of values and principles but does not push towards one specific government model.²⁵ For the UN, the “people” is both a source of legitimacy and the cornerstone of fundamental principles.²⁶ In 2002, the UN Commission on Human Rights declared a series of principles as essential to a democratic state. The list included among others: the respect of human rights and fundamental freedoms, freedom of expression and opinion, periodic free and fair elections by universal suffrage and secret ballot, a pluralistic system of political parties and organisations and free, independent and pluralistic media.²⁷ As long as a state respects and protects these rights, the actual governing regime is irrelevant for the UN. The question is more on what states do to ensure that the electoral process is protected. In this respect, the safeguards must not only aim at an efficient ballot day but must also protect the democratic debate both in social and traditional media and in public fora at large. Modern democracy is thus a system where the democratic debate has enlarged its forum. Ensuring free and fair elections has become difficult as safeguards must now be adapted to each circumstance.

²³ A. TOURAINE, “What is democracy?”, the UNESCO Courier, <<https://en.unesco.org/courier/novembre-1992/what-democracy>> last accessed 1 April 2021.

²⁴ M. RASHBROOKE, “3 ways to upgrade democracy in the 21st century”, TEDTALK, <https://www.ted.com/talks/max_rashbrooke_3_ways_to_upgrade_democracy_for_the_21st_century> last accessed 20 April 2021.

²⁵ Resolution 2002/46 from the Commission on Human Rights relating to further measures to promote and consolidate democracy, Office for the High Commissioner for Human Rights, <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjRoI-0m5XwAhXABmMBHbvQDw0QFjABegQIAxAD&url=http%3A%2F%2Fap.ohchr.org%2Fdocuments%2FE%2FCHR%2Fresolutions%2FE-CN_4-RES-2002-46.doc&usg=AOvVaw2ZKWqdaYAM699P8OwkVGel> last accessed 18 April 2021.

²⁶ UNITED NATIONS, Guidance notes from the Secretary-General on democracy, <https://www.un.org/democracyfund/sites/www.un.org.democracyfund/files/un_sg_guidance_note_on_democracy.pdf> last accessed 3 February 2021, p. 2.

²⁷ *The UN Commission on Human Rights is now the UN Human Rights Council.*

2.2. Free and fair elections: the voice of the people

Democratic elections involve many rights and liberties including the right to vote, the right to run for office, the right to free expression, the right to form and participate in political parties and the right to access to information.²⁸ It is the compliance with these fundamental rights and liberties that protects the integrity of the democratic process. But elections are not the sole attribute of democracy.²⁹ In fact, elections are also organised in non-democratic countries. As M. Ndulo and S. Lulo said, “*elections alone cannot be equated with democracy*” but it is an “*intrinsic component*”.³⁰ According to the United Nations, for an election to be free and fair, it must meet several requirements. Elections should be periodic, genuine, free, fair, done by universal and equal suffrage during a secret ballot and the results should be honestly and impartially counted.³¹ These criteria are intrinsically intertwined. Elections are meant to be the free expression of voters. For an election to be genuine it must allow information to trickle down the different layers of society and lead to an informed choice by the people. As a guarantee of freedom, the ballot is secret, and no voter should be compelled to disclose their political choice. Additionally, universal and equal suffrage means that all citizens have a right to vote and voice their opinion without any interference.

As elections decide who will be the decision-makers for a given period of time, they are essential moments for the community. Article 3 of Protocol no.1 to the European Convention on Human Rights specifies elections should be held at reasonable intervals. N. Urbinati considers elections as crucial within societies as they can also serve as a substitute to revolution.³² She wrote that “*elections simultaneously separate and link citizens and government. They create a gap between state and society at the same time as they allow them to communicate and even conflict, but never fuse. As the only institutional site of popular will and the point of contact between parliament and the extra-parliamentary realm, elections can be seen from different political angles*”.³³

²⁸ R. A. DAHL, “Democratic polities in Advanced Countries: Success and Challenge”, in *New Worldwide hegemony. Alternatives for change and social movements*, 2004, pp.51-70, <<http://biblioteca.clacso.edu.ar/ar/libros/hegeing/Dahl.pdf>> last accessed 30 May 2021, p.59.

²⁹ Y. LECUYER, *Le droit à des élections libres*, Conseil de l’Europe, « Hors Collection », Strasbourg, France, 2014, <<https://www-cairn-info.proxy.bib.ucl.ac.be:2443/le-droit-a-des-elections-libres--9789287177599.htm>>, p. 10.

³⁰ M. NDULO & S. LULO, “Free and fair elections, violence and conflict”, *Cornell Law Faculty Publications*, 51, 2010, <<https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=1185&context=facpub>> last accessed 3 March 2021, p.156.

³¹ OHCHR, Chapter 23: Monitoring Human Rights in the Context of Elections, in: OHCHR, *Manual on human rights monitoring*, 2011, Professional Training Series n°7, <<https://www.ohchr.org/Documents/Publications/Chapter23-MHRM.pdf>> last accessed 10 June 2021, p.8.

³² N. URBINATI, “Representative democracy and its critics”, in *The Future of Representative Democracy* (eds. S. ALONSO, J. KEANE, W. MERKEL), Cambridge University Press, 2011, p.2.

³³ N. URBINATI, *op. cit.* (voy. note 32), p.2.

Among the rudiments of conducting democratic elections are the principle of freedom and fairness. J. Elklit and P. Svensson referring to the work of R. Dahl defined the concepts of “free” and “fair”. Fairness is concept of impartiality. *“The opposite of fairness is unequal treatment of equals, whereby some people (or groups) are given unreasonable advantages. Thus, fairness involves both regularity (...) and reasonableness (...)”*.³⁴ Whereas *“freedom (...) contrasts with coercion. Freedom entails the right and the opportunity to choose one thing over another”*. Thus, all citizens should be able to vote without any influence, intimation or violence. Article 3 of Protocol No. 1 to the European Convention on Human Rights states: *“the High Contracting Parties undertake to hold free elections at reasonable intervals by secret ballot, under conditions which will ensure the free expression of the opinion of the people in the choice of the legislature”*.

Fairness and freedom are not “ballot-day only” principles. The principles of fairness and freedom must be observed before, on and after the polling day.³⁵ The results from a survey conducted by the European Parliament following the 2019 elections showed that the level of satisfaction with free and fair elections has increased in most of the EU member states since 2018. But citizens’ satisfaction cannot be enough. After the 2019 European elections, the European Commission Vice-President for Values and Transparency Vera Jourova said that *“we witnessed the most digital European election campaigns ever. This brings new opportunities of political engagement for the citizens and the candidates, but also pose some challenges for free and fair elections”*.³⁶ She feared voters could be subjected to unlawful manipulation techniques online. The Commission’s report on the elections stressed the need to protect democracy from manipulation and foreign interference.

2.3. The democratic crisis

2.3.1. The European integration

Modern western democracies saw the achievement of the ideas of the liberal thinkers of the 18th century.³⁷ From the 20th century onwards, liberal democracy became the political system associated with prospects of prosperity. But the World Wars put this aim on hold. World War II, even more than World War I, was an ideological war between states with liberal democracy regimes on the one

³⁴ J. ELKIT & P. SVENSSON, “The Rise of Election Monitoring: What makes elections free and fair?”, *Journal of democracy*, 8(3), 1997, p.35.

³⁵ J. ELKIT & P. SVENSSON, (voy. note 34), p.37.

³⁶ EUROPEAN COMMISSION, *Commission reports on 2019 European elections: fostering European debates and securing free and fair elections*, Press release, 2019, <https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1123> last accessed 1 June 2021, p. 1.

³⁷ R. LESAFFER, *European Legal History: a cultural and political perspective*, Cambridge, Cambridge University Press, 2009, p.496.

hand and states with fascist regimes on the other hand.³⁸ After WWII, the institutional process of the European integration started. Political decisions were made gradually on economic, social and political matters with the prime objectives of avoiding another war and of creating a liberal democratic front opposite the Soviet bloc. The European integration proceeded to systematic enlargements. It extended to the inclusion of central and eastern countries after the Cold War. The 2004 enlargement was perceived as a great achievement in pushing back the totalitarian ideology. However, seventeenth years later the optimism is somewhat fading away.

2.3.2. The rise of extreme parties in Europe

After the wave of democratisation, the 21st century saw the rise of populist and autocratic political streams. Extremist parties are gaining power throughout Europe and they are in a way challenging the existence of the Union all together. Right-wing populism and extremism do not intrinsically go hand in hand, though it is likely that extremist parties are also leaning towards populism.³⁹ Political scientist T. Greven mentions three reasons why right-wing populist parties fuel the rise of individualism and Euroscepticism. The first reason is that they create antagonists to the narrative. It is the “us against them” motto; “them” generally being minorities. He said that “*the more ethno-centric the conception of the people, the more xenophobic the positioning against “the other” and the clearer the desire to overthrow democratic governance, the more likely it is that a right-wing populist party is also extremist.*”⁴⁰ When we observe the case of Poland and its Eurosceptic, anti-immigration and anti-LGBTQAI President, Andrzej Duda, the “other” is any individual who does not meet the criteria to be part of the “us”. The relationship between Poland and the EU has been challenging for a few years now. The tension extended to the threat by the Union to remove EU funds if Poland kept violating the rule of law.⁴¹ But the threat had close to little impact. Despite the 2019 resolution on public discrimination and hate speech against LGBTI people, the towns which declared themselves “LGBT free zone” and were denied EU funds received funds from the Polish government instead.⁴²

³⁸ R. LESAFFER (voy. note 37).

³⁹ T. GOVERN, *The rise of right-wing populism in Europe and the United States: a comparative perspective*, Friedrich Ebert Stiftung, 2016, <http://dc.fes.de/fileadmin/user_upload/publications/RightwingPopulism.pdf> last accessed 13 April 2021, p.1-3.

⁴⁰ T. GOVERN, *op. cit.*, (voy. note 39), p.2.

⁴¹ A. TIDEY, “Poland to give money to “LGBT free” town denied EU funding”, Euronews, 2020, <<https://www.euronews.com/2020/08/19/poland-to-give-money-to-lgbt-free-towns-denied-eu-funding>> last accessed 13 April 2021.

⁴² A. TIDEY, *op. cit.* (voy. note 41).

The second reason is that populist parties resort to negativity in political communication whether it be by insulting individuals, by resorting to conspiracy theories, by propagating factually false narratives or breaking political correctness.⁴³ Andrzej Duda and the PiS party in Poland constantly referring to an “LGBT ideology” and comparing it to “*communism, Nazism and the plague*” has proven to be effective.⁴⁴ A survey realised in 2019 by OKO.press showed that almost half of the respondents over fifty consider the “LGBT ideology” as a threat to Poland.⁴⁵ The survey also highlights the generational division. The younger generation tends to disagree. They often represent the majority of the protestors defending human rights in the Polish streets.⁴⁶

Poland is one of many examples of the current cleaving climate in Europe. What has been demonstrated with Poland holds true in other EU member states. During the 2016 elections, populist parties gained or won across Europe (e.g. Netherlands, Austria, France). In Hungary, Viktor Orban has been re-elected for a third term. In Spain, the last parliament election saw the rise of far-right political formations. In parallel of the renewed interest for extreme parties, the interest in elections in general is challenged. In France a recent sociological study conducted by Ipsos and Ifop showed that the majority of the younger generation do not intend to vote in the next presidential election.⁴⁷ Beside this political apathy of the youth, the majority of the 25-34 years old who want to participate in the ballot are considering voting for the President of the far-right party *Rassemblement National*, Marine Le Pen. These parties are being elected and re-elected. It is no longer a one-time exception but a real movement on the rise. The question is thus how the EU can defend its values and remain relevant in the eyes of those who preferred to it nationalism and individualism.

⁴³ T. GOVERN, *op. cit.* (voy. note 39), pp. 1-3.

⁴⁴ S. WALKER, “Polish President issues campaign pledge to fight LGBT ideology”, The Guardian, 12 June 2020, <<https://www.theguardian.com/world/2020/jun/12/polish-president-issues-campaign-pledge-to-fight-lgbt-ideology>> last accessed 10 April 2021.

⁴⁵ A. AMBROZIAK, “Sondaż OKO.press. Homofobia nie działa. Większość Polaków odrzuca opowieść o groźnej ideologii”, OKO.press, 2020, <<https://oko.press/sondaz-wiekszosc-odrzuca-opowiesc-o-groznej-ideologii-lgbt/>> last accessed 13 April 2021.

⁴⁶ LE MONDE AVEC REUTERS, « En Pologne, des milliers de manifestants après l’arrestation d’une militante LBGT », Le Monde, 2020, <https://www.lemonde.fr/international/article/2020/08/09/en-pologne-des-milliers-des-manifestants-apres-l-arrestation-d-une-militante-lgbt_6048524_3210.html> last accessed 29 April 2021.

⁴⁷ F. JOHANNES, « Comment le Rassemblement national est devenu le premier parti de la génération des 25-34 ans », Le Monde, <https://www.lemonde.fr/politique/article/2021/04/05/le-rassemblement-national-premier-parti-des-25-34-ans_6075574_823448.html> last accessed 31 July 2021.

3. When the democratic debate occurs online

3.1. Social media and the democratic process

3.1.1. The revolution of Internet and social media

Time made the first purposes of the creation of the European Union obsolete. We do not fear a war in the traditional military sense anymore. The threat to our democracy is more insidious. And Internet offers tools that were not around in the 20th century. The digital era started out with huge democratic expectations. Indeed, as digital sociologist D. Cardon explains: Internet broadened the forum where the public debate takes place.⁴⁸ Cardon pinpointed two revolutionary elements of Internet. First, Internet allows any citizen to voice his or her opinion. And in democratic societies, these opinions should be expressed freely. Second, the notion of “public” and “private” sphere is increasingly blurred.⁴⁹ Certain discussions occurring previously behind closed doors are now happening through tweets, comments and posts. Opinions of experts and professionals are not the only voices to be heard anymore.

Social media has had its positive impact on certain democratic upsurges. In 2011, online communication helped organised part of the Arab Spring protests.⁵⁰ Internet and social media offered a window to go around official and government-approved media. Even though certain commentators suggests that the western lenses tended to exaggerate the impact of social media in the Arab uprisings, the protestors did resort to some extent to social media for their organisation.⁵¹ Social media platforms helped these revolutions but as A. Touraine said: “*revolution sweep away an old order: they do not create democracy*”.⁵² Internet is a mean to an end but cannot stand alone in this process. The digital transformation is certainly not the first challenge democracy has been facing. Certain authors talk about a “representation crisis” whereas others doubt that we even reached that point. M. A. Cohendet connects the representation crisis to the growth of private interests in decision making.⁵³ She considers that citizens have better access to information. They are thus more incline to critic the representatives and their alleged renunciation in the face of private interests. According to Cohendet, this ineffectiveness or failure of representation creates an atmosphere of disaffection and mistrust.

⁴⁸ D. CARDON, *La démocratie Internet : promesses et limites*, Paris, Seuil, 2010, p.11.

⁴⁹ D. CARDON, *op. cit.* (voy. note 48), p. 36.

⁵⁰ V. DE COOREBYTER, *Des pratiques démocratiques de crise*, Les @nalyses du CRISP en ligne, 2015, <https://www.crisp.be/crisp/wp-content/uploads/analyses/2015-09-17_ACL-de_Coorebyter_V-2015-Des_pratiques_democratiques_de_crise.pdf> last accessed 3 May 2021, p. 4.

⁵¹ M. MANRIQUE, « Réseaux sociaux et médias d'information », *L'Harmattan « Confluence Méditerranée »*, 79, 2011, <<https://www.cairn.info/revue-confluences-mediterranee-2011-4-page-81.htm>> last accessed 13 accessed 2021, p.84

⁵² A. TOURAINE, *op. cit.* (voy. note 23).

⁵³ M. A. COHENDET, « Une crise de la représentation politique », *Presses Universitaires de France*, 2004/2 (18), pp. 41-61, <<https://www.cairn.info/revue-cites-2004-2-page-41.htm>> last accessed 30 July 2021, pp. 48-51.

However, the philosopher and political scientist B. Manin defends that the two indicators of a representation crisis are not met.⁵⁴ It requires: the decrease of voters' participation rates and the discredit of politicians. If the second is debatable the first is unquestionable at the European level: people still go to the ballot. Participation rates are even going up. If we look at the last European elections, with a global European turn out of 50.66% it is the highest results since 1994.⁵⁵ The European Parliament has commissioned an extensive analysis of the 2019 elections turnout.⁵⁶ Regarding the question whether their country benefited from the EU, the survey shows that respondents from Poland (-13%) and Finland and Lithuania (both -12%) were less incline to consider that the EU contributed to positive outputs for their country. However, by looking at the turnout from these member states it did not stop the citizens from voting. In fact, they all have improved their turnouts since 2014. Thus, people who disagree with the EU or are even Eurosceptic made their way to the ballot. Whether or not we reached a representative crisis, democracy can be in crisis as a result of different factors such as the lack of transparency, the decrease of accountability, the disaffection with politics or the increase of extreme votes.⁵⁷

3.1.2. Materializing the virtual debate

Internet often offers a platform for social change. The broadcasting of personal experiences of social inequalities can spark a necessary democratic debate on social subjects. It helps raise awareness. The #MeToo movement and the Black Lives Matter movement took off online and only after reached traditional media and politics. Internet is a powerhouse when it comes to bringing forward fundamental social issues. However, to have a real impact, these movements had to somehow be materialized outside of the virtual world. Belgian philosopher and political scientist V. de Coorebyter considers that there are limits to any virtual debate.⁵⁸ By looking at movements such as Black Lives Matter in the United States, the Yellow Vests in France or the Global Climate March, they all resulted from an anger that grew online but then gained more legitimacy when they reached the streets. Conversely, if connecting people was originally at the root of internet and social media platforms, it has now evolved and reached new purposes. Journalist and activist Carole Cadwalladr stressed the fact that technology tends to divide individuals and even disrupt the electoral process.⁵⁹ The

⁵⁴ H. LANDEMORE, *op. cit.* (voy. note 17), p. 10.

⁵⁵ J. ZALC, N. BECUWE & A. BURUIN, *The 2019 post-electoral survey: have European elections entered a new dimension, Eurobarometer Survey 91.5 of the European Parliament*, A Public Opinion Monitoring Study, 2019, p.19.

⁵⁶ J. ZALC, N. BECUWE & A. BURUIN, *op. cit.* (voy. note 55).

⁵⁷ S. A. ERCAN & J-P. GAGNON, "The Crisis of Democracy: Which Crisis? Which Democracy?", *Democratic Theory*, 1(2), 2014, pp.1-10.

⁵⁸ V. DE COOREBYTER, *op. cit.* (voy. note 50), p. 4.

⁵⁹ C. CADWALLADR, *op. cit.* (voy. note 3).

democratic advantages of Internet during the electoral period can be outshined by the rallying power of demagoguery online.

3.2. The social and algorithmic phenomena online

3.2.1. Capitalising on online engagement

The algorithms on social platforms purposely show users the content that he or she will likely react to.⁶⁰ As profit-oriented companies, platforms capitalize on the engagement of the user. When a user reads, clicks, likes, shares or reacts to a content, he is confronted with advertisement. The longer the user stays engaged with a certain content, the more likely it is that he will be recommended similar content in the future. Based on users' data, the algorithm will prioritize videos, images or posts in order to maximise profit. However, what the algorithm does not do is determining whether the content shared is true or false. It focuses only on users' engagement. As a result, online platforms are by design predisposed to spreading disinformation. What is striking is that statistically false information spreads more easily online than true information. Online platforms act as a megaphone, amplifying the fake. A study realised by MIT scholars S. Vosoughi, D. Roy and S. Aral on 126,000 cascades of tweets,⁶¹ showed that false information was 70% more likely to be re-tweeted than the truth. The study highlighted that one of the factors leading to sharing was the novelty of the information. Moreover, any interaction with the content sends a signal to the algorithm to display it to other users.⁶² Finally, the important takeaway of this study for this master thesis is that false political news enjoys the most effective propagation.

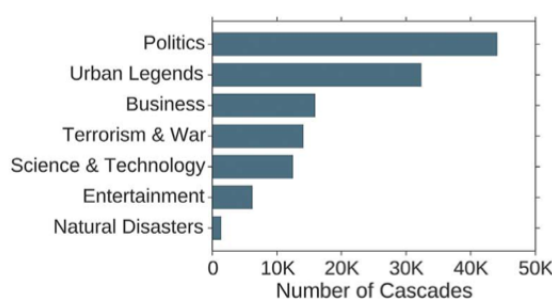


Figure 1 - Histogram of the total number of rumours cascade in their data across the seven most topical categories; from SOROUGH VOSOUGHI, DEB ROY & SINAN ARAL, *the spread of true and false news online*

⁶⁰ N.I. BROWN, *op. cit.* (voy. note 5), p.19.

⁶¹ S. VOSOUGHI, D. ROY & S. ARAL, *The spread of true and false news online*, MIT initiative on the digital economy, 2018, pp. 1146-1151, <https://ide.mit.edu/sites/default/files/publications/2017%20IDE%20Research%20Brief%20False%20News.pdf> last accessed 2 April 2021, pp.1146-1149.

⁶² S. VOROUSHI, D. ROY & S. ARAL, *op. cit.* (voy. note 61), pp.1146-1149.

3.2.2. Human behaviour online

The question is why do people keep believing what they read online despite ongoing warnings? Several phenomena may explain part of the reason why the spread of false or manipulated information is occurring and even rising. In our digital era, human beings are constantly confronted with a flow of information. They have thus accepted that they cannot be omniscient on every news.⁶³ Their behaviour is in fact a coping mechanism. An information cascade for example is a behaviour consisting in relying on other people's knowledge. D. Easley and J. Kleinberg in their book about connectiveness in modern society consider that a cascade is developing "*when people abandon their own information in favour of inferences based on earlier people's actions*".⁶⁴ It is the idea that if the majority believes or does something, it must probably be the right thing to believe or do. This is the case even though this might indeed contradict what one knows or believes to know. When one is flooded with information, there is comfort in conformity.

People online are also influenced by the way information is presented. Negative information tends to be better remembered than positive information.⁶⁵ People even crave for negative news stories and are more inclined to spread them.⁶⁶ And when these kinds of news are confirming people's beliefs, it fuels human beings' habit to surround themselves with confirmative information.⁶⁷ This phenomenon is called the "filter bubbles". And as already demonstrated, platforms' algorithms also encourage them. Online, we are filled with posts from our friends, with content shared by our acquaintances or with videos from pages we do not even follow. To remain economically viable, platforms want you to spend time reading, scrolling, "liking" and commenting. To do so, the algorithm presents content that will appeal to the user. And yes, that implies providing content confirming what he or she consider to be true. Everyone's Facebook, Twitter, Instagram is different and specifically tailored to respond to the expectations of its owner. It creates a bubble of self-criticism isolation where the rule is no longer "the truth" but "my truth". Researchers have demonstrated that victims of the disinformation mostly do not even read debunking articles.⁶⁸

⁶³ B. CHESNEY AND D. CITRON, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 107, 2019, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954#> last accessed 3 July 2021, pp.1765-1766.

⁶⁴ D. EASLEY AND J. KLEINBERG, *Networks, Crows, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, 2010, <<https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf>> last accessed 18 July 2021, p.484. p.484.

⁶⁵ E.A. KENSINGER, "Negative emotions enhances memory: Behavioural and Neuroimaging Evidence", *Sage Journal*, 16(4), pp.213-218, 2007, p.218.

⁶⁶ B. CHESNEY et D. CITRON, *op. cit.* (voy. note 63), p.1766.

⁶⁷ B. CHESNEY et D. CITRON, *op. cit.* (voy. note 63), p.1768.

⁶⁸ B. CHESNEY et D. CITRON, *op. cit.* (voy. note 63), p.1768.

Disinformation especially when conveying negative and novel news, sticks longer than any other kind of news.

American jurist Cass R. Sunstein also talked about a similar concept to filter bubbles: echo chambers. It is just another way of saying that the same idea, belief or information is repeated in a closed circle where other opinions cannot easily interfere.⁶⁹ It is also in these bubbles or chambers that conspiracy theories thrive. Faced with these concurring phenomena, Cass R. Sunstein and American economists M. Gentzkow and J. M. Shapiro refer to an ‘*ideological segregation*’ online. Meanwhile, French sociologist P. Flichy prefers to refer to a “*fracture démocratique*” (a democratic divide). In both cases, the result is the same: the current functioning of social media intensifies sub-groups within society without exposing the said sub-groups to opposing beliefs.

3.3. Intermediate conclusion

The democratic process of elections is effective when the population is provided with accurate information. In theory, Internet offers the democratic opportunity both to provide information to a broader portion of the population and to confront them with different viewpoints. It could make every citizen an enlightened elector. But like the *Magistracies* of the Ancient Greek democracy, candidates have understood that lack of information and disinformation can help manipulate the results of an election. A misinformed voter is a voter that can potentially shift its opinion when presented an idea in a certain light. Currently, the way in which Internet and social media work offers many openings to interfere with elections. And finding the adequate solution against it requires to first understand in practice the risks our societies are facing.

4. Manipulation of elections: disinformation and confusion

4.1. Disinformation and the right to free and fair elections

Disinformation impacts the integrity of the elections process. It blurs the line between truth and fiction, amplifies voters’ confusion and obstructs a fact-based political debate.⁷⁰ The campaign leading up to the polling day is a time when citizens can be familiarised with candidates’

⁶⁹ P. MASSA, “Trust It Forward: Tyranny of the Majority or Echo Chambers?” In *The Reputation Society: how online opinions are reshaping the offline world* (eds. H. MASUM & M. TOVEY), Cambridge, MIT Press, 2012, p. 27.

⁷⁰ NATIONAL DEMOCRATIC INSTITUTE, “Disinformation and Electoral Integrity: A Guidance Document for NDI Elections Programs”, 2019, <https://www.ndi.org/sites/default/files/Disinformation%20and%20Electoral%20Integrity_NDI_External_Updated%20May%202019%20%281%29.pdf> last accessed 10 June 2021.

programmes, ambitions and personalities. Having access to accurate information is an essential part of the election process. In April 2019, Switzerland's Supreme Court overturned a nationwide referendum about tax law on the grounds that the citizens were not given enough information to make an enlightened decision.⁷¹ The Court said the lack of information and the lack of transparency violated the "*freedom of the vote*".⁷² Sufficient information is part of a respectable democratic process. Thus, disinformation challenges the requirements of a fair and free electoral process. The U.S.-based NGO, National Democratic Institute (hereinafter the "NDI"), lists in its guidance document on disinformation and electoral integrity a number of "*globally recognized standards*" for democratic elections: the right to seek, receive and impart information in order to make an informed choice on elections day, the right to a level playing field among the candidates and the freedom of opinion, expression and the press.

4.2. Freedom of expression and the access to accurate information

Each of the AI-driven election manipulation systems that this paper will analyse have one common denominator: disinformation. When an election is around the corner, the press assumes part of the responsibility to inform and report on the candidates and their political agenda. And voters should be able to create their own opinions on the candidates and their political projects based on accurate and true information. The right to freedom of expression and information is enshrined in article 11 EU Charter of Fundamental Rights. The right to hold an opinion also derives from the same article. The free and informed expression of people's opinion is one of the prerequisites to free and fair elections (see 2.2. free and fair elections: the voice of the people).

The rights to freedom of expression and freedom of opinion are essential human rights standards part of a free and fair electoral process.⁷³ In *Bowman v. The United Kingdom*, the European Court of Human rights stated the following: "*Free elections and freedom of expression, particularly freedom of political debate, together form the bedrock of any democratic system. (...) For this reason, it is particularly important in the period preceding an election that opinions and information of all kinds are permitted to circulate freely.*"⁷⁴ Thus, the freedom to communicate on issues and ideas about the

⁷¹ J. HENLEY, "Court overturns referendums as voters were poorly informed... in Switzerland", The Guardian, 2019, <<https://www.theguardian.com/world/2019/apr/11/switzerland-court-overturns-referendum-as-voters-were-poorly-informed>> last accessed 12 June 2021.

⁷² I. NENADIC, *op.cit.* (voy. note 2), p.2.

⁷³ OHCHR, Chapter 23: Monitoring Human Rights in the context of elections, in *Manual on human rights monitoring*, Professional Training Series n°7, Geneva, 2011 <<https://www.ohchr.org/documents/publications/chapter23-mhrm.pdf>> last accessed 3 June 2021, pp.9-10.

⁷⁴ ECHR, *Bowman v. The United Kingdom*, 19 February 1998, §42.

public and political situation is fundamental. Freedom of expression also implies the right to equal access to information, guaranteed media pluralism and the existence of a free media environment. The idea behind media plurality is that voters should “*receive diverse range of information to enable them to make an informed choice*”.⁷⁵

However, the right to freedom of expression and opinion can be restricted. Limits are allowed if they are in conformity with the law, necessary in a democratic society and based on specific grounds (e.g. national security, territorial integrity or public safety, the prevention of disorder or crime, the protection of health or morals, the protection of the reputation or rights of others, the prevention the disclosure of information received in confidence).⁷⁶ The public debate must be free, but such freedoms should be balanced with other collective interests. When conflicting with other fundamental rights (e.g. equality between candidates, right to access information, right to personal integrity...), restrictions are possible but must be proportionate to the legitimate aim pursued. Restriction will for example encompass situation of defamation, hate speech or incitation to violence. The objective is to maintain a healthy democratic process with access to accurate information. But the challenges posed by the spread disinformation online questions the efficacy of existing safeguards.

4.3. Disinformation disorders

4.3.1. The notion of disinformation

Internet deeply changed the media landscape and the way people consume information. It has also multiplied sources of information. If we know where to look, we can find on the Internet all essential tools to partake and engage in the democratic debate. In 2012, the European Court of Human Rights established: “(...) *the Internet has now become one of the principal means by which individuals exercise their right to freedom of expression and information, providing as it does essential tools for participation in activities and discussions concerning political issues and issues of general interest.*”⁷⁷ This positive aspect shows its shortcomings when measured against the current atmosphere online including the tendency towards online disinformation and “fake news”. While these two terms often go hand in hand, they do not equate to the same thing. Fake news is an umbrella term that does not have a stable meaning, but which includes all types of information that do not

⁷⁵ OHCHR, *op. cit.* (voy. note 73), p.12.

⁷⁶ European Charter of Fundamental Rights, art. 11 §2.

⁷⁷ Cour. Eur. D.H., *Ahmet Yildirim c. Turkey*, 18 December 2012, req. n° 3111/10, § 54.

meet the threshold of trustworthiness and verifiability of a reliable news.^{78 79} An information can be false (*misinformation*), harmful (“*malinformation*”) or both (*disinformation*).⁸⁰ Computer scientists K. Shu and S. Wang and information and technology scientists D. Lee and H. Liu add three other disinformation disorders to the list⁸¹. *Rumours*, meaning unverified and doubtful information that, when they are confirmed to be false, become a type of misinformation; *urban legend* which are generally false stories linked to local popular culture and *spam*, or messages sent to a large scale of individuals containing unwanted “*irrelevant or inappropriate information*”. But these manifestations can easily be classified within the three main categories of misinformation, “malinformation” and disinformation.

4.3.2. Disinformation, not fake news

The term “fake news” is already problematic. In the introduction to his book *The Reality Game*, media and technology researcher Samuel Woolley expressly excludes the use of the term ‘fake news’. He explains his choice by a simple argument: fake news became itself a tool to spread fake news.⁸² The philosopher and epistemologist, Joshua Habgood-Coote even argues that the term ‘fake news’ poses a threat to democracy as it has been weaponised to manipulate people into supporting certain political aims.⁸³ The word alone provokes a reaction and triggers public emotions. We can hear Donald Trump shouting his well-known: ‘You are Fake News!’. No statistics, no facts, no counterarguments. The term is also a strong ally for propaganda since it has no clear definition. It leaves room for any individual to apply its own interpretation to it. Through repetition and its often-dramatic narrative, the term gained significance in the collective psyche without having a distinctive meaning. Given the topic of this research and even though the term ‘fake news’ is still used by some

⁷⁸ UNESCO, *Journalist, Fake news and Disinformation: Handbook for Journalism, Education and Training*, UNESCO Series on Journalism Education, 2018, p.44-56.

⁷⁹ J. HABGOOD-COOTE, “Stop talking about fake news!”, *Inquiry*, 62 (9-10), 2018, pp.1033-1065, <<https://www.tandfonline.com/doi/abs/10.1080/0020174X.2018.1508363?journalCode=sinq20>> last accessed 2 February 2021, pp. 1036-1037.

⁸⁰ C. WARDLE & H. DERAQSHAN, *Information Disorder: Towards an interdisciplinary framework for research and policy making*, Council of Europe Report, 2017, p.20.

⁸¹ K. SHU, S. WANG, D. LEE, H. LIU et. al., “Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements”, in *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* (eds. K. SHU, S. WANG & H. LIU), Springer International Publishing, 2020, p.3.

⁸² S. WOOLLEY, *The Reality Game: How the next wave of technology will break the truth and what can we do about it*, London, Endeavour, 2020, p. IX.

⁸³ J. HABGOOD-COOTE, *op. cit.* (voy. note 79), pp. 1036-1037.

academic researchers, this paper will not refer to ‘fake news’, preferring instead the well-defined term ‘disinformation’.⁸⁴

4.3.3. The rationales of disinformation

a. Disinformation through the years

Disinformation deliberately creates confusion, and it may even attempt to manipulate the opinion of people, organization and states or seek to harm them.⁸⁵ Disinformation is not a modern disease. There are examples of disinformation in Ancient Rome already.⁸⁶ Octavian for example started a propaganda campaign targeting Mark Anthony’s reputation. The purpose was to gain Roman citizens’ good graces by peddling false information about his opponent. False information included Anthony being a drunk and a politician disregarding traditional Roman values. Disinformation was already instrumentalised in the fight for power between the two authoritative political figures, successors of Julius Caesar.

In the mid-15th century, the printing press invented by Gutenberg drastically changed the way news circulated throughout Europe.⁸⁷ After years of orally shared or slowly written down by hand information, this innovation allowed the mass distribution of news and gossip. Written documents were used for the propaganda of the states and the church. It led for example to the publication of *Malleus Maleficarum* (Hammer of Witches) a treatise on how to hunt, prosecute and execute witches. What these examples demonstrate is that one of the recurrent results of disinformation is polarisation and division. P. Baines and N. Snow wrote that “*the objective of disinformation is not necessarily to persuade but to sow division, to engender doubt, to foster acquiescence; and in all of this, actual belief does not necessarily play any role at all*”.⁸⁸ If some examples of disinformation can sound ridiculous, the impact it has differs from one group to another. Among many factors for such

⁸⁴ See examples: N. M. BRASHIER, G. PENNYCOOK, A. J. BERINSKY & D. G. RAND, “Timing matters when correcting fake news”, *Proceedings of the National Academy of Sciences of the United States of America*, 118(5), <<https://www.pnas.org/content/118/5/e2020043118>> last accessed 30 July 2021; A. KAPLAN, “Chapter 9: Artificial intelligence, social, media, and fake news: is this the end of democracy?”, in *Digital Transformation in Media & Society* (eds. A. A. GÜL et al.), Istanbul University Press, 2020, <<https://www.pnas.org/content/118/5/e2020043118>> last accessed 30 July 2021, pp.149 et s.

⁸⁵ C. WARDLE & H. DERAKSHAN, *Information Disorder: Towards an interdisciplinary framework for research and policy making*, Council of Europe Report, 2017, p.20.

⁸⁶ J. POSSETTI & A. MATTHEWS, *A short guide to the history of ‘fake news’ and disinformation: a learning module of journalists and journalism educators*, International Center for Journalists, 2018, <https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf> last accessed 28 February 2021, p. 2.

⁸⁷ J. POSSETTI & A. MATTHEWS, *op. cit.* (voy. note 86), pp. 1-2.

⁸⁸ P. BAINES & N. SNOW, *The SAGE handbook of propaganda*, SAGE publications Ltd., 2020, p. 55.

difference, the European Commission mentions the “*education level, democratic culture, trust in institutions, the inclusiveness of electoral systems, the role of money in political processes, and social and economic inequalities*”.⁸⁹ Dr. I. Nenadic adds to the above elements “*the status of journalism and legacy media, levels of media literacy, the political context and legal safeguards*”.⁹⁰ Because it leverages such objective factors in order to achieve the greatest possible impact on its intended audience, targeted disinformation can be highly effective. It feeds off personal information to determine where disinformation will have greater effect. In 2016, Cambridge Analytica selected its targeted audience based on personal data points to understand how to influence their electoral behaviour (see section 5.2).

b. Disinformation disorder: symptom of other diseases

Disinformation is thus as old as news itself. Innovations such as the printing press but also more recently the Internet and social networks have been important boosters of disinformation. However, the spread of disinformation is symptomatic of other economic, technological, political and ideological issues.⁹¹ The European Commission in its approach to tackle online disinformation pinpointed three main triggers. Firstly, disinformation is a symptom of our rapidly changing modern societies facing many challenges including “*economic insecurity, rising extremism, and cultural shifts*”, which all are “*breeding ground for disinformation campaigns to foster societal tensions, polarisation, and distrust*”.⁹² Disinformation and the current European democratic crisis (see section 2.3) are encouraging one another in a vicious circle. And the algorithmic and social phenomena online can nurture disinformed beliefs through cognitive and confirmation bias (see section 3.2).

Secondly, the European Commission links disinformation with challenges currently facing the information media. The evolution our societies undergoes also impacts the media sector. The fact that online platforms have become active in news reporting has had an overall negative impact on the media industry. Media had to rethink their business models to be economically viable. According to the European Commission, the need to be paid by the number of clicks and reading time led to more quantitative over qualitative content, making it harder to distinguish qualitative articles within the mass.

⁸⁹ COMMISSION EUROPENNE, *Tackling Online Disinformation: an European Approach*, Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, 2018, p.4.

⁹⁰ I. NENADIC, *op.cit.* (voy. note 2), p.2.

⁹¹ COMMISSION EUROPENNE, *op.cit.* (voy. note 89), p. 4.

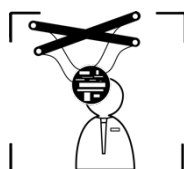
⁹² COMMISSION EUROPENNE, *op.cit.* (voy. note 89), p. 4.

Thirdly, social media have all the ingredients to spread disinformation. Disinformation is amplified simply by sharing decisions of the users themselves. As we have seen, the more an information is false, the more likely it is that it will be shared online.⁹³ Platforms like Twitter are often used as a primary source of information, especially for current events. During the 2015 Paris terrorist attacks, social media offered immediate communication and conveyed information directly from people on site. It became a source of information among users. Social media are the “media of the instant” where we share information first and verify the news afterwards.

5. Artificial intelligence: electoral manipulation through disinformation

By spreading inaccurate information, some artificial intelligence tools can help fuel confusion among voters. Resorting to AI to disinform citizens during the democratic electoral process is becoming a weapon of election manipulation.⁹⁴ Before analysing some of the EU responses to these phenomena, the following section will explain how AI (sub-section 5.1. deepfakes, sub-section 5.2. targeted influence campaigns, sub-section 5.3. political bots) may impede free and fair elections by encouraging false narratives.

5.1. Deepfake: a future tool or an overrated threat?



Let us imagine a scenario where elections are upcoming. A few days prior to poll day, a video of a candidate starts spreading online. In this viral video, the candidate is saying offensive words and appears to be intoxicated at his workplace.

5.1.1. Deepfake in action

The first AI-driven disinformation mechanisms are deepfakes. Deepfake is a portmanteau word melding “deep learning”, a form of artificial intelligence derived from machine learning, and “fake”.⁹⁵ Deepfake algorithms can create fake audio, pictures or videos. They can for example overlay images on a source person to create the illusion of the person saying or doing something that he or she never said or did.⁹⁶ The term “deepfake” emerged in 2017 when a user of the social platform

⁹³ S. VOSOUGHI, D. ROY & S. ARAL, *op.cit.* (voy. note 61), pp.1146-1151.

⁹⁴ N.I. BROWN, “Deepfakes and the Weaponization of Disinformation”, *Virginia Journal of Law & Technology*, 23(1), 2020, <https://static1.squarespace.com/static/5e793709295d7b60295b2d29/t/5f0b7e4e0f4f095f7e70613f/1594588750542/v23il_1-Brown.pdf> last accessed 31 July 2021, pp. 7-8.

⁹⁵ T.L. WAGNER ET A. BLEWER, “The Word Real is No Longer Real: Deepfakes, Gender and the Challenges of AI-Altered Video”, *Open Information Science*, 3(1), 2019, p.36.

⁹⁶ T.T. NGUYEN et. al., *Deep Learning for Deepfakes Creation and Detection*, Cornell University, 2019, <<https://arxiv.org/pdf/1909.11573.pdf>> last accessed 13 February 2021, p. 1.

Reddit resorted to open-source face-swapping technology on pornographic videos under the username “Deepfakes”.⁹⁷ The first deepfakes were thus video manipulations where the faces of adult film stars were replaced with faces of other women, especially famous actresses. The fakes were obviously realised without the subjects’ consent.

If deepfake has been pioneered on content from the porn industry, it has since then been used for other purposes. The term “deepfake” now encompasses more broadly all AI-generated manipulated images or videos. The technology can, for example, recreate the voice of a CEO on the phone committing money theft⁹⁸ or manipulate the speech of a Prime Minister connecting the pandemic to climate change.⁹⁹ It can display politicians cursing and calling names a U.S. President or show someone engaging in criminally reprehensible behaviours.¹⁰⁰ The options to manipulate reality are manifold. R. Chesney and D. Citron depict it as a new generation of disinformation.¹⁰¹

5.1.2. The technology behind the fake

The most common way of creating a convincing deepfake is to resort to Generative Adversarial Networks (GAN). Created by Ian Goodfellow and his colleagues in 2014, the GAN approach uses two algorithms working against each other to obtain a realistic output.¹⁰² The adversarial process works in two steps: the generator neural network (NN Generator) generates new data (e.g. a face) and the discriminator (NN Discriminator) evaluates the authenticity of the data (e.g. determining if the face it picked is real or fake). It is often compared to a game of cop and robber because, after rounds of this exercise, the NN Generator will get better at generating data while the NN Discriminator will get better at recognising fakes. It provides a good example of machine learning: the system learns by training and constantly adjusting the result, either with or without human guidance, in order to capture the characteristics of the person and obtain an increasingly believable result.¹⁰³

⁹⁷ T.L. WAGNER ET A. BLEWER, *op. cit.* (voy. note 95), p.33.

⁹⁸ J. DAMIANI, “A Voice deepfake was used to Scam a CEO out of \$243,000”, <<https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=d85064622416>> last accessed 22 February 2021.

⁹⁹ EXTINCTION REBELLION, « La vérité sur le Covid-19 et la Crise écologie : un discours par Sophie Wilmès », <<https://www.extinctionrebellion.be/fr/dire-la-verite>> last accessed 20 February 2021.

¹⁰⁰ K. FAGAN, “A viral video that appeared to show Obama calling Trump a ‘dips—’ shows a disturbing new trend called ‘deepfakes,’” Insider, 2018, <<https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4?r=US&IR=T>> last accessed 22 February 2021.

¹⁰¹ R. CHESNEY & D. CITRON, *op. cit.* (voy. note 6).

¹⁰² I. J. GOODFELLOW et al., *Generative Adversarial Nets*, Neural Information Processing Systems Conference Paper, 2014.

¹⁰³ I. J. GOODFELLOW et al., *op. cit.* (voy. note 102).

5.1.3. Deepfakes and elections manipulation: the ongoing warning

a. *The skills required to create a deepfake*

One of the essential components for the realisation of a deepfake are images of the targeted subject. Depending on the type of fake, one or many images will be needed. The more images the neural network will have, the more accurate and believable the results of its training will be. And again, the Internet and social media play a role by making readily available multiple images of the same person. It is no longer difficult to obtain the basic material to develop the deepfake. Deepfake are also often presented as accessible and increasingly easy to create. In its Communication on disinformation, the European Commission described deepfake as a “*new, affordable, easy-to-use technology*”.¹⁰⁴ But it might be a stretch. If photos are easily accessible, the creation of a deepfake based on them is not within everyone’s skills. It requires an understanding of the technology and of the software.

However, images manipulation did not wait for machine learning and sophisticated artificial intelligence to attempt to mirror reality.¹⁰⁵ What we are often confronted with today are cheap fakes or shallow fakes. The latter are manipulated videos implying less complex technical tools and software than deepfakes.¹⁰⁶ But the results are less convincing. Imperfections include glitches, motion slowing down or speeding up erratically, cut or unnatural movements of the subjects. All are ‘tells’ that the video footage is not authentic. In *Deepfake and Cheap Fakes*, B. Paris and J. Donovan stressed the fact that deepfake and cheap fake are two ends of one spectrum.¹⁰⁷ According to Paris and Donovan, the evolution from one end to another depends on the technology behind the fake. Thus, if a face swap using a real time filter application is a cheap fake rather easy to produce, a voice synthesis or a virtual performance using the GAN approach is a deepfake requiring a certain level of expertise.

¹⁰⁴ COMMISSION EUROPÉENNE, *Tackling Online Disinformation : a European Approach*, Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, 2018, p.5.

N.I. BROWN, “Deepfakes and the Weaponization of Disinformation”, *Virginia Journal of Law & Technology*, 23(1), 2020, <https://static1.squarespace.com/static/5e793709295d7b60295b2d29/t/5f0b7e4e0f4f095f7e70613f/1594588750542/v23i1_1-Brown.pdf> last accessed 31 July 2021, p.6.

¹⁰⁵ B. PARIS & J. DONOVAN, *Deepfakes and Cheap Fakes: the Manipulation of Audio Visual Evidence*, Data & Society, 2020, <https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf> last accessed 10 March 2021, pp.11-12.

¹⁰⁶ R. GOLDZWEIG & M. BRADY, *Deepfakes: How prepared are we? Multi-stakeholder perspectives and a recommendations roadmap*, Democracy Reporting International, 2020, <<https://democracy-reporting.org/wp-content/uploads/2020/11/2020-11-Deepfakes-Publication-No-2-Web-file-1.pdf>> last accessed 10 March 2021, p.10.

¹⁰⁷ See the cheap fakes and deepfakes spectrum: PARIS B. & DONOVAN J., *op. cit.* (voy. note 105), pp. 11-12.

The General Adversarial Approach of Ian Goodfellow is thus one of the many ways to create an audio-visual manipulation. It will result in a highly realistic fake, but the technology is not accessible to just any curious or malicious mind. Computer scientists are the most skilled to obtain a believable result. Nevertheless, there are other deepfakes techniques for people with minimum computer knowledge and time on their hands.¹⁰⁸ A face swap can for example be done on Adobe After Effects. The idea of deepfake being easy to achieve can therefore be true depending on the type of fake under consideration. The most believable deepfakes are still restricted to a certain range of skilled individuals. For now, at least.

b. Deepfake: real threat to elections?

Alongside the general disinformation issue, deepfakes poses a threat to candidates' image, reputation and credibility. Even though there are possible remedial actions against the fakes, such as legal action for defamation, there are limits to it. First, a deepfake as any online content is difficult to make disappear. It can be replicated, shared, downloaded multiple times. The removal of the harmful content is hardly ever permanent. Secondly, a shared content knows no boundaries. It can be difficult to find the perpetrators. If a deepfake can become viral within a few hours, the effort required to take it down and obtain reparation can take years.¹⁰⁹

But could deepfakes really swing an election? Or could it have a significant impact on the results? Prior to the 2020 U.S. elections, the risk was broadly announced. The fear was that deepfakes would be the equivalent of the "fake news" phenomenon of the 2016 elections.¹¹⁰ During a hearing in June 2019, the House Permanent Select Committee on Intelligence discussed the deepfakes' threat to national security and democratic institutions.¹¹¹ The same concerns are now raised regarding the 2022 French elections.¹¹² However, in practice the projections of the involvement of these AI-

¹⁰⁸ B. PARIS & J. DONOVAN, *op. cit.* (voy. note 105).

¹⁰⁹ N. MATIN, "Online predators spread fake porn of me: here's how I fought back", TEDTalk, TED Talk, 2017, <https://www.ted.com/talks/noelle_martin_online_predators_spread_fake_porn_of_me_here_s_how_i_fought_back> last accessed 5 May 2021.

¹¹⁰ SPECIAL COUNSEL ROBERT S. MUELLER III, *Report on the Investigation into Russian Interference in the 2016 Elections*, 1(1-5), 2019, <<https://www.justice.gov/archives/sco/file/1373816/download>> last accessed 30 July 2021.

¹¹¹ HOUSE INTELLIGENCE COMMITTEE, *National Security Challenges of Artificial Intelligence, Manipulated Media, and Deepfakes*, 116th Congress, 2019, <<https://www.congress.gov/event/116th-congress/house-event/109620>> last accessed 3 June 2021.

¹¹² L. RONFAUT, « "Deepfakes" : avant la présidentielle, le grand faux dans l'inconnu », Libération, 2021, <https://www.liberation.fr/economie/economie-numerique/deepfakes-avant-la-presidentielle-le-grand-faux-dans-linconnu-20210514_ZDSIMQUA5BH3TNJBC2AJ5I4O3E/> last accessed 5 July 2021.

generated videos during the 2020 US elections were never met.¹¹³ The risk remained hypothetical. The longstanding warning probably shielded the elections from AI-form of disinformation.

Only a few instances of political deepfakes exist and many are just satirical. In 2018, the sp.a (Flemish socialist party) posted on its social media a deepfake of Trump encouraging Belgium to withdraw from the Paris agreement.¹¹⁴ The purpose was to “*start a public debate*” and “*draw attention to the necessity to act on climate change*”, said the sp.a spokesperson to Politico.¹¹⁵ The objective was not to confuse the viewers. The relatively low quality of the video, the stiffness of the gestures and the muppet-like movement of the mouth left little doubt about its fake nature. However, when looking at the comments under the post, it appeared that some viewers had misunderstood the irony of the message. Even without the intent to deceive, deepfakes can lead to deception and doubt especially when the audience is not educated enough about the mere existence of deepfakes.

But maybe the reason we cannot find many political deepfakes examples is because they are getting harder to detect. A video of Ali Bongo Ondimba, the President of Gabon who made no public appearance for several months, raised questions in December 2018.¹¹⁶ Experts have looked at the video without being able to determine whether it was a deepfake or not. Deepfake experts A. Ovadya and H. Farid said that aside from the difficulty of assessing the nature of the video, the simple possibility of a video being a deepfake is already damaging. It is not the deepfake alone that poses a threat. The circumstances in which it is shared matter. Within the European member states, the ongoing climate of mistrust and Euroscepticism can foster a “good” narrative for those willing to influence European elections.

5.1.4. Deepfakes and fundamental rights

a. The right to information and the Liar's Dividend

The interplay between disinformation and free and fair elections has already been discussed (see point 4.1). As a form of disinformation, the same dynamics hold true for deepfakes. What

¹¹³ T. SIMONITE, “What Happened to the Deepfake Threat to the Election?”, Wired, 2020, <<https://www.wired.com/story/what-happened-deepfake-threat-election/>> last accessed 3 June 2021.

¹¹⁴ See the video posted on Twitter here: <https://twitter.com/vooruit_nu/status/998089909369016325>; see the video posted on Facebook here: <<https://www.facebook.com/watch/?v=10155618434657151>>.

¹¹⁵ H. VON DER BURCHARD, “Belgian socialist party circulated ‘deepfake’ Donald Trump video”, Politico, 2018, <<https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/>> last accessed 14 June 2021.

¹¹⁶ A. BRELAND, “The Bizarre and Terrifying Case of the “Deepfake” Video that helped bring an African Nation to the Brink”, Mother Jones, 15 March 2019, <<https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/>> last accessed 16 June 2021.

differentiates deepfakes from other types of disinformation is that it is an AI-created fake image or video. The fairness of an election is also evaluated based on the transparency of its process.¹¹⁷ Information allows public scrutiny.¹¹⁸ During the 2020 US campaign, Kathryn Harrison, founder and CEO of the DeepTrust Alliance said she feared more a late true video impossible to determine as being true or false than any actual deepfake. This phenomenon known as the so-called “Liar’s Dividend” is a concept in which “*people with a propensity to deceive are given the benefit of an environment in which it is increasingly difficult for the public to determine what is true*”.¹¹⁹ Bobby Chesney and Dannielle Citron who coined the term, consider that deepfakes facilitate an easy way out of accountability for politicians and candidates.¹²⁰ Video evidence can be dodged by the argument that we cannot trust our own eyes anymore. Paradoxically, the mere existence of deepfakes thus offer a chance of a credible denial even in the face of strong evidence of misconduct or of an outright lie.

b. The right to non-discrimination and freedom of press

Deepfakes: when the main targets are women

As a disinformation weapon and beside the right to freedom of information *per se*, deepfakes jeopardise other rights. According to a report from Deeptrace, women are the first victims of deepfakes.¹²¹ 96% of deepfakes created nowadays are porn deepfakes.¹²² And the entirety of porn deepfakes are targeting women. On the other side, political deepfakes mostly target men. From this distinction, the link between porn deepfake and elections manipulation is not self-explanatory. It is indeed more insidious. Porn deepfakes intend to threaten, objectify or harass women. The psychological and physical impact for the victims is vilifying. Victims may suffer from panic attacks, job loss and they may even attempt to commit suicide.¹²³ An Australian victim known under the pseudonym Noëlle Martin said the porn deepfake of herself shared online impacted her education,

¹¹⁷ J. ELKLIT & P. SVENSSON, “The Rise of Election Monitoring: What makes elections free and fair?”, *Journal of democracy*, 8(3), 1997, p.37.

¹¹⁸ UNITED NATIONS, *Freedom of Expression and Elections in the Digital Age, Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, 2019, Research Paper 1/2019, p.5.

¹¹⁹ HOUSE INTELLIGENCE COMMITTEE, intervention of D. CITRON, National Security Challenges of Artificial Intelligence, Manipulated Media, and Deepfakes, 116th Congress, 2019, <<https://www.congress.gov/event/116th-congress/house-event/109620>> last accessed 3 June 2021.

¹²⁰ B. CHESNEY & D. CITRON, *op. cit.* (voy. note 60), p.1785.

¹²¹ H. AJDER, G. PATRINI, F. CAVALLI & L. CULLEN, *The State of Deepfakes: Landscape, Threats and Impact*, Rapport Deeptrace, 2019, pp.1-2.

¹²² H. AJDER et al., *op. cit.* (voy. note 121), p.1.

¹²³ S. MADDOCKS, “‘A Deepfake Porn Plot Intended to Silence Me’: exploring continuities between pornographic and ‘political’ deep fakes”, *Porn Studies*, 7(4), 2020, DOI: 10.1080/23268743.2020.1757499, pp. 415-423.

employability, her personal relationships and her mental health.¹²⁴ The deepfake produces a multi-layered attack. And the harassment does not stop with the suppression of the video. The video may be downloaded, shared on other platforms, commented on social media and the victim can be persecuted and threatened. On top of that, the video exists in the collective mind, whether it has been demonstrated to be fake or not.

The case of women journalists and activists

The impact of a deepfake for the reputation of one person can be far-reaching. But when the targets are women journalists or activists, the impact can become democratic. In 2018, the Indian investigation journalist and author Rana Ayyub, was the victim of a porn deepfake. She received death threats, misogynistic comments and she was constantly cyber-bullied.¹²⁵ The reason was partly political. India was in a period of unrest following the rape and murder of an 8-year-old Kashmiri. And the ruling nationalist party organised marches in support of the accused. Rana Ayyub took a stand against the actions of the party which launched a smear campaign against her. She described it herself as “*a plot intended to silence [her]*”.¹²⁶ As a journalist, an activist and a person perceived as “*anti-establishment*”, the campaign undeniably aimed at suppressing her voice from the public debate. In her own words, these events changed her: “*From the day the video was published, I have not been the same person. I used to be very opinionated, now I’m much more cautious about what I post online. I’ve self-censored quite a bit out of necessity*”.¹²⁷ The situation required the intervention of the United Nations for her government to finally ensure the protection of Rana Ayyub’s life.¹²⁸ According to digital rights advocate S. Maddocks, there is a form of continuity between porn deepfakes and political deepfakes.¹²⁹ As women are the main deepfakes targets, AI-generated videos are just another tool to silence the voices and opinions of part of the population. As these deepfakes target journalists, this also goes against the right to freedom of press and the principle of media pluralism (art. 10 ECHR). States should therefore ensure that media are free, independent and

¹²⁴ L.CHUNG, ““Lifelong impacts’: image-based abuse skyrockets online amid pandemic”, 2020, <<https://www.smh.com.au/national/nsw/lifelong-impacts-image-based-abuse-skyrockets-online-amid-pandemic-20201201-p56jpu.html>> last accessed 3 May 2021.

¹²⁵ R. AYYUB, *I was the Victim of a Deepfake Porn Plot Intended to Silence Me*, 2018, <https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316> last accessed 8 May 2021).

¹²⁶ R. AYYUB, *op. cit.* (voy. note 125).

¹²⁷ R. AYYUB, *op. cit.* (voy. note 125).

¹²⁸ UNITED NATIONS HUMAN RIGHTS OFFICE, *UN experts call on India to protect journalist Rana Ayyub from online hate campaign*, 2018, <<https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23126&LangID=E>> last accessed 25 April 2021.

¹²⁹ S. MADDOCKS, *op. cit.* (voy. note 123), pp.415-423.

diverse. By protecting media diversity, democratic states have the responsibility to allow different and opposing standpoints in the press.¹³⁰

A multi-shaped weapon of discrimination

The example of Rana Ayyub demonstrates the different form the discreditation and undermining of certain voices can take (disinformation campaign, deepfakes, cyber-harassment...). The right to freedom of opinion and expression is a “*central pillar of democratic societies, and a guarantor of free and fair electoral processes, and meaningful and representatives public and political discourse*”.¹³¹ The 2014 UN report on the promotion and the protection of the right to freedom of opinion and expression also stressed states’ responsibility to “prohibit” incitement of discrimination and violence, especially during electoral periods. The report showed concerns regarding hostility targeting religious minorities, members of the LGBT+ community, women and journalists.¹³² Deepfakes are exactly the type of weapon that can be used to silence women or discourage them from becoming public figures. Poet and writer H. Mort, whose face was edited on various violent sexual images in 2020,¹³³ said she was: “*punished for being a woman with a public voice of any kind*”.¹³⁴ In 2020, the UNESCO conducted a survey on online violence against women journalists (901 respondents from 25 to 50 years old) which revealed that image-based abuse was already experienced by 15% of the respondents while only 4% mentioned having been victims of deepfakes.¹³⁵ If it not yet the main means of intimidation against women journalists, 73% of the women respondents did say having experienced online violence in connection with their work.¹³⁶ Other than the mental health impact, the outcomes of the harassment according to the reports are manifold: the withdrawal of the journalist for a period of time (17%) or permanently (11%) from participating in online conversations, the resignation from their job (4%) or the end of their journalistic practice (2%).¹³⁷

¹³⁰ UNITED NATIONS, *op. cit.* (voy. note 128), p.5.

¹³¹ UNITED NATIONS, *Report of the Special Rapporteur on the promotion and the protection of the right to freedom of opinion and expression*, Frank La Rue, General Assembly, Human Rights Council, 2014, p.5.

¹³² UNITED NATIONS, *op. cit.* (voy. note 131), p.3.

¹³³ S. ROYLE, “Deepfake porn images still give me nightmares”, BBC, 2021, <<https://www.bbc.com/news/technology-55546372>> last accessed 4 July 2021.

¹³⁴ K. HAO, *Deepfake porn is ruining women’s lives. Now the law may finally ban it*, MIT Technology Review, 2021, <<https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/>> last accessed 29 June 2021.

¹³⁵ J. POSETTI et al., *Online violence Against Women Journalists: A global Snapshot of Incidence and Impacts*, UNESCO, 2020, <<https://www.icfj.org/sites/default/files/2020-12/UNESCO%20Online%20Violence%20Against%20Women%20Journalists%20-%20A%20Global%20Snapshot%20Dec9pm.pdf>> last accessed 15 June 2021, p.7.

¹³⁶ J. POSETTI et al., *op. cit.* (voy. note 135), p. 5.

¹³⁷ J. POSETTI et al., *op. cit.* (voy. note 135), p.11.

Article 21 European Charter of Fundamental Rights and article 14 ECHR prohibit any discrimination based on grounds including sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation. Historically, elections are deeply linked to the principle of equality. The notion of equality has always been advocated within democracies. It is among its core principles. Even at the time of the Athenian democracy where women were excluded from the poll, equality between free men was advocated (see point 2.1.1). The elections are meant to represent the opinion from “the people” as an entity encompassing the sum of its equal parts. Discriminatory behaviours should not impede the equal flow of ideas necessary to the democratic electoral process. Resorting to discrimination to remove citizens from the debate will ultimately distort the elections’ results. And when it comes to women journalists, it also concerns the protection (or lack thereof) of the freedom of press (art. 10 ECHR).

5.2. Data collection and AI in political advertising



For the upcoming elections, some voters were hesitating between the different candidates at first but by the time the deepfake of one of the candidates became viral their mind was made up. His main opponent seemed way more ready for the job. For several months now, the undecided voters have seen content demonstrating that the candidate was a drunk unfit for office.

5.2.1. The Cambridge Analytica Scandal

The second AI-driven disinformation mechanisms to analyse are targeted influence campaigns developed through data collection. By browsing the Internet, we constantly agree, knowingly or not, to provide personal or non-personal information about ourselves. It can be information about our habits, leisure, values, beliefs but also our social environment, family status or political views. These data are intrinsically connected to our identity. Data has thus now become a highly valuable asset for companies seeking to gain economic benefits, but also for political campaigns aiming at influencing elections results.¹³⁸

Cambridge Analytica has been one of the biggest election manipulation scandals in recent history. It has shaken the United States in the wake of the 2016 elections. In March 2018, the

¹³⁸ J. NOUJAIM & K. AMER, “The Great Hack”, Documentary, 2019.

Facebook/Cambridge Analytica dossier was brought to the public's attention by *The New York Times*, *The Observer of London* and *The Guardian*.¹³⁹ It revealed a scheme using personal data to engineer personalised political advertising on a massive scale. Cambridge Analytica was an American-own data-driven company based in London and specialising in communications strategy. The company claimed to have 5,000 data points on over 220 million Americans that could help classify and predict the personality and behavioural patterns of U.S. voters. Cambridge Analytica collected data through an online personality quiz shared on Facebook. The quiz called "ThisIsYourDigitalLife" was created by researcher Aleksandr Kogan with whom Cambridge Analytica had a partnership.¹⁴⁰ The free personality quiz was in fact profiling respondents.

Accessing the quiz also meant consenting to granting access to the respondents' Facebook data and to their friends' liked pages.¹⁴¹ The data collection thus extended to information from the respondents' profiles (e.g. Facebook "likes" and shares or pages followed), as well as from their friends' profiles. Indeed, another factor compounding the extent of the data harvest was the lack of data protection afforded by Facebook itself. The social network gave extended access to the personal information of users who did not give consent for their data to be collected. In total, the quiz is known to have collected data of at least 87 million Facebook users.¹⁴²

The collected data helped classify American personalities into different categories with the goal of conducting a microtargeted influence campaign. The test was able to determine the psychological profile of thousands of people. These people were evaluated on five personality traits (the so-called "Big Five"): openness, conscientiousness, extroversion, agreeableness and neuroticism (also called OCEAN scores).¹⁴³ This personality profile combined with various information on personal preferences harvested from the digital footprints lead to rather accurate voters' profiles. Thus, the combination of behavioural psychology and profiling with big data allowed for algorithms to microtarget the intended audience with influential content. As Cambridge Analytica CEO Alexander

¹³⁹ N. CONFESSORE, "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far", *The New York Times* 2018, <<https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>> last accessed 5 July 2021.

¹⁴⁰ C. CADWALLADR & E. GRAHAM-HARRISON, "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach", *The Guardian*, 2018, <<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>> last accessed 5 July 2021.

¹⁴¹ R. RATHI, *Effect of Cambridge Analytica's Facebook ads on the 2016 US Presidential Elections*, towards data science, 2019, <<https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us-presidential-election-dacb5462155d>> last accessed 24 April 2021.

¹⁴² C. CADWALLADR & E. GRAHAM-HARRISON, *op. cit.* (voy. note 140).

¹⁴³ R.J. GONZALEZ, "Hacking the citizenry? Personality profiling, 'big data' and the election of Donald Trump", *Anthropology Today*, 33(3), 2017, <https://doi.org/10.1111/1467-8322.12348>, p.10.

Nix explained: “*psychographic [provides] an understanding of your personality (...) it is personality that drives behaviour and behaviour that (...) influences how you vote*”.¹⁴⁴

The focus of that campaign was on what Cambridge Analytica called “*the persuadables*”; the voters whose opinion was not set on a particular candidate and could be influenced into voting for Cambridge Analytica’s end client: Donald Trump. The campaign thus personally targeted around 20 million “*persuadable*” voters, especially in swing states.¹⁴⁵ Cambridge Analytica’s strategy combined three elements (see Figure 2 below) to achieve behavioural change. The algorithmic method was to send personalised ads, videos, “memes”, blog articles to the “*persuadables*” in order to set their mind into voting for the republican candidate.

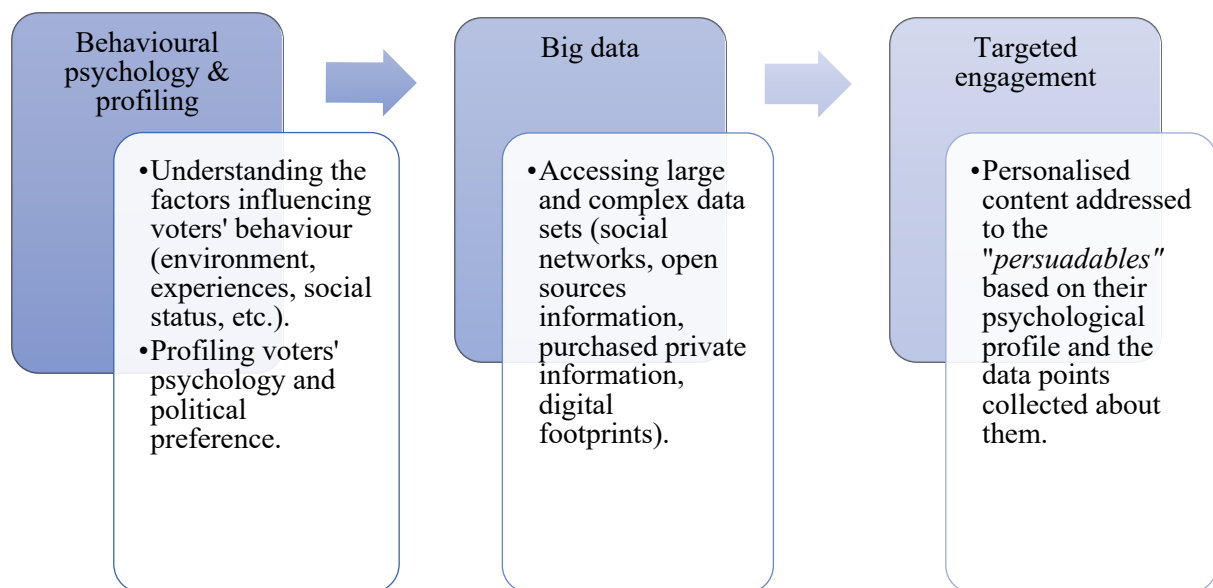


Figure 2 – Cambridge Analytica’s strategy based on three elements: behavioural psychology, big data and targeted engagement

Cambridge Analytica created an algorithmic mechanism able to influence elections. The tool created by the company has even been described by certain authors as a “*weaponized AI propaganda machine*”.¹⁴⁶ The automated propaganda algorithm is “*targeting people individually to recruit them to an idea. (...) They’re capturing people and then keeping them on an emotional leash and never*

¹⁴⁴ A. NIX, “The Power of Big Data and Psychographics, Conference during the 2016 Concordia Annual Summit”, see the video here: <<https://www.youtube.com/watch?v=n8Dd5aVXLCc>> last accessed 17 July 2021.

¹⁴⁵ R.J. GONZALEZ, *op. cit.* (voy. note 143), p.9.

¹⁴⁶ B. ANDERSON & B. HORVATH, “The rise of the weaponized AI propaganda machine”, Scout, 2017, <<https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b>> last accessed 9 July 2021.

letting them go”, said professor and data journalist Jonathan Albright.¹⁴⁷ The complexity was also that the origin of the political ads, montages or memes were unknown. They were shared online without being directly linked to any candidate’s campaign.

*“Sometimes you could use proxy organizations who are already there, you feed them with the material (...) and they do the work. We just put information into the bloodstream of the internet and then, (...) watch it grow, give it a little push ever now and again over time, to watch it take shape. And so, this stuff infiltrates the online community and expands, but with no branding so its unattributable, untrackable.”*¹⁴⁸

Mark Turnbull, Managing Director of Cambridge Analytica

5.2.2. The impact of Cambridge Analytica

Just as any disinformation campaign, it remains highly difficult to evaluate and quantify the actual impact of the tactics on the results. It is even more complicated when the microtargeting campaign occurred in such opacity. In fact, prior to the moment former Cambridge Analytica employee Christopher Wylie became a whistleblower, little was known about the innovative approach of advertisement by the company.¹⁴⁹ The appeal of saying that Donald Trump won the elections thanks to AI intervention and complex profiling is a shortcut that will not be taken here. In fact, social scientists Chen and Potenza considered that a simple personality test where respondents could misrepresent themselves and potentially old Facebook likes could not constitute enough data to create accurate political profiles.¹⁵⁰ Donald Trump’s victory was multifactorial. Excluding Cambridge Analytica altogether is not reasonable but pinpointing the specific impact of the company is not possible either. However, Trump did win six states won by Barack Obama in 2012 by a very short difference of votes with rival presidential candidate, Hillary Clinton: Michigan, Wisconsin, Iowa,

¹⁴⁷ C. CADWALLADR, “Google, democracy and the truth about internet search”, The Guardian, 2016, <<https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook>> last accessed 30 June 2021.

¹⁴⁸ CHANNEL 4 NEWS INVESTIGATIONS TEAM, “Exposed: Undercover secrets of Trump’s data firm”, Channel 4 News, 2018, <<https://www.channel4.com/news/exposed-undercover-secrets-of-donald-trump-data-firm-cambridge-analytica>> last accessed 9 July 2021.

¹⁴⁹ B. EKDALE & M. TULLY, “African Elections as a Testing Ground: Comparing Coverage of Cambridge Analytica in Nigerian and Kenyan Newspapers”, *African Journalism studies*, 40(4), 2019, pp.27-43, DOI: 10.1080/23743670.2019.1679208, p.28.

¹⁵⁰ A. CHEN & A. POTENZA, “Cambridge Analytica’s Facebook Data Abuse shouldn’t get credit for Trump”, The Verge, 2018, <<https://www.theverge.com/2018/3/20/17138854/cambridge-analytica-facebook-data-trump-campaign-psychographic-microtargeting>> last accessed 3 July 2021.

Pennsylvania, Ohio and Florida.¹⁵¹ Following the scandal, Cambridge Analytica filed for bankruptcy in 2018.¹⁵²

5.2.3. A global issue

a. *Digital colonialism*

The implication of Cambridge Analytica has been exposed following its involvement in the 2016 U.S. elections. And if their work for the Trump campaign has been the most talked about, it is not their first electoral interference. They worked for example for an electoral campaign in Trinidad and Tobago where young black citizens were influenced not to vote to benefit the United National Congress party (the “UNC”), a mostly Indian party who won the 2010 elections.¹⁵³ Cambridge Analytica also recognised working on elections campaign in Malaysia, Kenya, Ghana, India, Romania to name a few. What is striking is that most of the states concerned were developing countries. Cambridge Analytica claimed to have tested their methods before moving to developed countries. Researcher and political analyst N. Nyabola calls these interferences in foreign democratic process “*digital colonialism*”.¹⁵⁴ Digital colonialism “*extracts profit but invests nothing in the country*”.¹⁵⁵ Experimenting in these specific countries at an earlier stage had obvious benefits for Cambridge Analytica. First, these countries had less or no privacy regulations. Cambridge Analytica could therefore develop its algorithm with more freedom and impunity. Second, these types of campaigns work best in locations where latent tensions exist. Political breeding grounds may include for example the fear of a terrorist group like Boko Haram in Nigeria or existing ethnic divisions in Kenya or Trinidad and Tobago.¹⁵⁶ The campaigns tapped into those national issues to create a cascade of information feeding off the emotions of the voters and providing them with one solution: electing the designated candidate.

¹⁵¹ R.J. GONZALEZ, *op. cit.* (voy. note 143), p.9.

¹⁵² O. SOLON & O. LAUGHLAND, “Cambridge Analytica closing Facebook data harvesting scandal”, The Guardian, 2018, <<https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-closing-down-after-facebook-row-reports-say>> last accessed 4 July 2021.

¹⁵³ “*We were working for the Indians; we went to the client and we said we want to target the youth and we want to try and increase apathy. The campaign had to be non-political, because the kids don't care about politics, it had to be reactive because they're lazy. So, we came up with this campaign that was all about being part of the gang, doing something cool, being part of a movement, and it was called the “Do So” campaign. Do So, don't vote. It's a sign of resistance against not the government but politics and voting. We knew that when it came to voting all the afro Caribbean kids wouldn't vote, because they Do So, but all the Indian kids would do what their parents told them to which, is to go out and vote... The difference in 18- to 35-year-old turnout was like 40% and that swung the election about 6%, which was all we needed in an election that's very close.*” Alexander Nix, CEO of Cambridge Analytica in J. NOUJAIM & K. AMER, “The Great Hack”, 2019.

¹⁵⁴ B. EKDALE & M. TULLY, *op. cit.* (voy. note 149), p.30.

¹⁵⁵ N. NYABOLA, *Digital Democracy, Analogue Politics: How the Internet Era is Transforming Politics in Kenya*, London, ZED Books, 2018, p.200.

¹⁵⁶ B. EKDALE & M. TULLY, *op. cit.* (voy. note 149), p.29 & p.37.

b. Brexit and the Leave.EU campaign

In Europe, Cambridge Analytica has also allegedly been involved in the Brexit referendum. Whistleblower Christopher Wiley testified that Cambridge Analytica was involved in the Brexit vote results through AggregateIQ, a company linked to the American communications strategy company. The Director of Program Development of Cambridge Analytica Brittany Kaiser can also be seen on videos at the launch of Leave.EU¹⁵⁷, answering questions regarding the communication strategy. However, Cambridge Analytica denies having done paid or unpaid work for the Leave campaign. Political campaigning is regulated in the UK. The amount of money allowed to be used on political campaign is legally limited. And in this case, suspicions extended to the financial resources that were collected and used through different entities to conduct the Leave campaign. The investment may have exceeded the authorized amount. But all of this remains hard to prove. Just as for assessing the impact of Cambridge Analytica on the 2016 U.S. elections. Who saw the political adverts on their feeds? Based on what personal data? How much money was paid for the adverts to be broadcasted? By whom was it even paid? And were there any foreign interventions? So many questions that the lack of transparency make unanswerable. The opacity and non-collaborative behaviour of Facebook is highlighting the problem of the algorithmic black box. In addition, Facebook CEO Mark Zuckerberg refused on several occasions to participate to the International Grand Committee meeting investigating the issue and discussing solutions.¹⁵⁸ He also refused to provide evidence regarding the allegations.

5.2.4. A weapons-grade communication technique

Contrary to Marc Zuckerberg, former director and whistleblower Brittany Kaiser did accept to discuss and give evidence to the UK. During her intervention in front of the House of Commons culture committee, she referred to Cambridge Analytica's strategy as "*a weapons-grade communication technique*".¹⁵⁹ Cambridge Analytica parent company SCL Group was also

¹⁵⁸ DIGITAL, CULTURE, MEDIA AND SPORT COMMITTEE OF THE HOUSE OF COMMONS, "Disinformation and 'fake news': Final Report", Eighth Report of Session 2017-19, 2019, <<https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf>> last accessed 7 July 2021, p.9.

¹⁵⁹ DIGITAL, CULTURE, MEDIA AND SPORT COMMITTEE OF THE HOUSE OF COMMONS, Oral evidence: Fake News, questions to the witness Brittany Kaiser, former Director of Program Development at Cambridge Analytica, 2018, video of the meeting: <<https://www.parliamentlive.tv/Event/Index/e5ae6255-c88e-4e62-bbf4-9c0c18ba7b6b>> last accessed 11 July 2021. Written evidence of the meeting: <<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/oral/81592.pdf>> last accessed 11 July 2021.

specialised in psychological warfare in a military context.¹⁶⁰ Far from a mere data-driven communications strategy start-up, its daughter was a company with an history of psyops using military strategy on uninformed citizens. Propaganda is defined as “*information, ideas, opinions, or images, often only giving one part of an argument, that are broadcast, published, or in some other way spread with the intention of influencing people’s opinions*”.¹⁶¹ Cambridge Analytica’s strategy is a new form of propaganda. A comparison between “traditional propaganda” and computational propaganda can help highlight the threat and the new legal challenges it poses (see Figure 3 below).

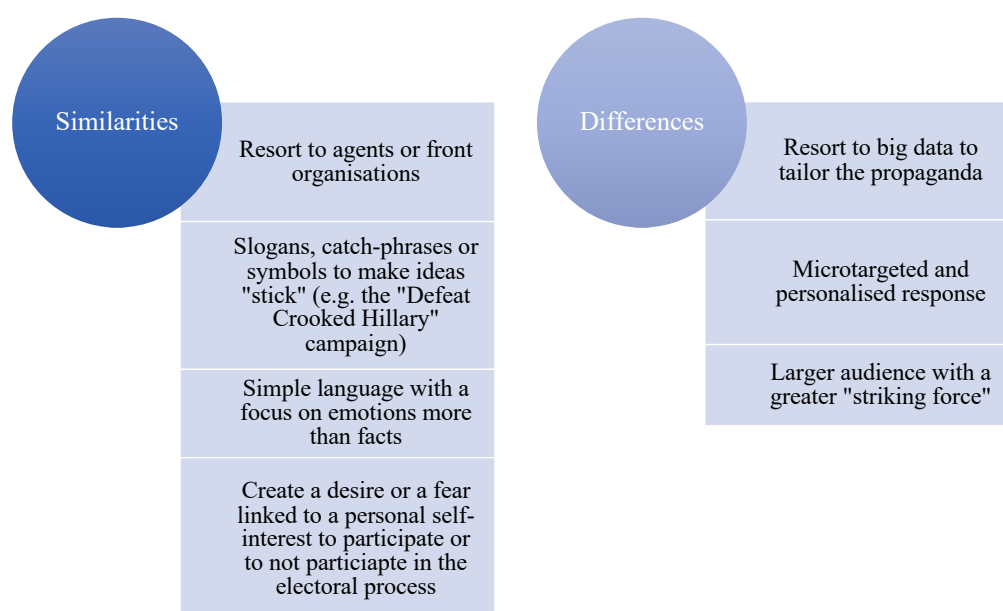


Figure 3 – Similarities and differences between “traditional” propaganda and micro-targeted political propaganda

The Cambridge Analytica technology uses traditional propaganda techniques, but it can reach even more people online. The slogans and catch-phrases put in montages, “memes” or short videos are bombarded to the target audience. And due to the way in which an algorithm like Facebook’s works, it will only be shown to users who could potentially positively respond to the political stimuli. The data harvested facilitates the microtargeting of those individuals. And the technology even allowed for the personalisation of the broadcasted content. As a result, part of the population has probably never even seen these political ads. But those who were indeed targeted were fed with a simple

¹⁶⁰ C. CADWALLADR, “The great British Brexit robbery: how our democracy was hijacked”, The Observer, 7 May 2017, full article on <<http://davelevy.info/wp-content/uploads/2015/03/great-british-brexite-robbery.pdf>> last accessed 1 July 2021.

¹⁶¹ CAMBRIDGE DICTIONNARY, <<https://dictionary.cambridge.org/fr/dictionnaire/anglais/propaganda>> last accessed 18 June 2021.

equation of active propaganda and disinformation: “we have a problem you should fear, here is the solution or behaviour to adopt to solve the issue”. The technique is thus to re-route the attention from political facts to personal emotional responses.

The fact that this process occurs online in complete opacity also raises questions as to the safeguards we have against foreign interventions. Free and fair elections are also elections conducted without the intervention of foreign interests. The 448-pages Mueller report in the United States said that Russian interference during the 2016 US elections was “*sweeping and systemic*”.¹⁶² Among the key findings of the report were the proof of attacks through information warfare on social media, hacking of Hillary Clinton’s campaign documents and emails and the release of stolen materials through entities including Wikileaks.¹⁶³ Foreign intervention also undermines democracy by fraudulently influencing the results and interfering in domestic affairs. But again, proving foreign interventions is difficult as it can take so many different shapes.

5.2.5. Targeted influenced campaign and fundamental rights

a. Right to privacy

Targeted political propaganda harms the right to privacy, enshrined in article 12 Universal Declaration of Human Rights, article 8 European Convention of Human Rights (hereinafter: ECHR) and article 7 European Charter of Fundamental Rights (hereinafter: Charter). The right to privacy is also linked to the notions of autonomy. As J.-J Rousseau explained in *Du Contrat Social*, a representative democracy is intended to elect those who will work for the common welfare of the society. However, within that society, every individual shall be able to keep its free will. Article 9 ECHR and article 10 Charter mention the right to freedom of thought. Representatives shall be elected by citizens who voted freely and autonomously on who will represent them. However, is a manipulative campaign enough to trigger a violation of a right to privacy and autonomy? Manipulating someone means interfering with one’s ability to critically reflect on a choice and influencing one’s decision-making without their awareness.¹⁶⁴ In that sense, microtargeting campaigning harms the sense of autonomy by reducing the plurality of the democratic debate. The

¹⁶² SPECIAL COUNCEL ROBERT S. MUELLER III, “Report on the Investigation into Russian Interference in the 2016 Elections”, 1(1-5), 2019, < <https://www.justice.gov/archives/sco/file/1373816/download>> last accessed 30 July 2021, p.1.

¹⁶³ SPECIAL COUNCEL ROBERT S. MUELLER III, *op. cit.* (voy. note 162), pp.14-35, pp.36-50, pp.50-51.

¹⁶⁴ D. SUSSER, B. ROESSLER & H. NISSENBAUM, “Technology, autonomy, and manipulation”, Alexander von Humboldt Institute for Internet and Society, *Internet Policy Review*, 8(2), Berlin, 2009, pp. 1-22, <<http://dx.doi.org/10.14763/2019.2.1410>> last accessed 13 July 2021, p.8.

harm is both individual and collective. The electoral process being an important democratic moment within society, the principle of autonomy is in even higher need to be safeguarded. As it “*undermines the core values*” of democracy, online manipulation is an ethical, social and political problem.¹⁶⁵

b. Right to data protection and the principle of transparency

As data is the primary fuel of the technology, targeted political propaganda also threatens data protection. The notion of private life of article 8 Charter and article 16 TFEU has been extended by case law to encompass the current development of new technologies.¹⁶⁶ The European Court of Human Rights defined personal data based on article 2 of Convention 108 as “*any information relating to an identified or identifiable individuals*”.¹⁶⁷¹⁶⁸ It is important to note that not all uses of data in politics is illegitimate. Nowadays, campaigning highly relies on data. Legitimate uses are traceable and resort to data that are “*public, lawfully obtained and at least partly anonymized*”.¹⁶⁹ The damage is when the use of data is illegitimate. And as AI is developed around data, “*the quality of a datasets used to build a model will directly influence the outcomes it produces*”.¹⁷⁰ It concerns both personal and non-personal data. As T. Timan and Z. Mann wrote, issues relating to big data analytics are twofold: it can help with monitoring and profiling and it can obtain personal data and/or sensitive information based on accessible non-personal data.¹⁷¹¹⁷² Privacy and data protection are among the “freedoms” of the EU Charter: freedom of thoughts, conscience and religion (article 10) freedom of expression and information (article 11) and freedom of assembly and association (article

¹⁶⁵ D. SUSSER, B. ROESSLER & H. NISSENBAUM, (voy. note 164), p. 11.

¹⁶⁶ ECHR, *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, 27 June 2017, §133.

ECHR, Guide to the Case-Law of the European Court of Human Rights, Data Protection, 2020, <<https://rm.coe.int/guide-data-protection-eng-1-2789-7576-0899-v-1/1680a20af0>> last accessed 20 July 2021.

¹⁶⁷ Convention no. 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data of 1 October 1981, art. 2.

¹⁶⁸ ECHR, *Haralambie v. Roumanie*, 27 October 2009, §77.

¹⁶⁹ K. MANHEM & L. KAPLAN, “Artificial Intelligence to Privacy and Democracy”, *The Yale Journal of Law and Technology*, 21, <https://yjolt.org/sites/default/files/21_yale_j.l._tech._106_0.pdf> last accessed 2 August 2021, p.138.

¹⁷⁰ S. HOLLAND et. al., “The Dataset Nutrition Label: a Framework to Drive Higher Data Quality Standards”, in *Data Protection and Privacy: Data Protection and Democracy* (eds. D. HALLINAN et al.), Hart Publishing, 2020, <<https://media.bloomsburyprofessional.com/rep/files/9781509932740sample.pdf>> last accessed 19 July 2021, p. 1.

¹⁷¹ T. TIMAN & Z. MANN, “Data Protection in the Era of Artificial Intelligence: Trends, Existing Solutions and Recommendations for Privacy-Preserving Technologies”, in *The Elements of Big Data Value* (eds. E. CURRY et al.), Springer International Publishing, 2021, p 156.

¹⁷² *Sensitive data are a special category of personal data which can reveal someone’s political beliefs, sexual orientation, race or ethnic origin, biometric or genetic information, religious or philosophical beliefs. Special rules apply to sensitive data (e.g. explicit consent or disclosure of information of utmost importance to meet other legal requirements).*

12). All these EU Charter articles could be endangered by the facilitated information dissemination procured by a data breach.¹⁷³

Within EU law, data can only be processed fairly and for specified purposes. It must be (i) lawful (ii) be done based on the subject's consent or any other legitimate interest (iii) and be necessary in a democratic society.¹⁷⁴ Everyone also has the right to access data collected on their behalf and the right to have it rectified. If the fundamental right to data protection exists, in practice the opacity of the technology makes it difficult to even be aware of an actual or potential breach. The way in which the data is collected and the extent of the data harvested are generally unknown to the public.

Thus, targeted political campaigning also threatens ethical political campaigning and the principle of transparency. AI specialists such as F. Pasquale refer to the “black box phenomenon” as the moment an algorithm has so much opacity and the variables are so complex that even computational experts struggle to understand it.¹⁷⁵ The information haze concerns the functioning of the algorithm, the data used and the output. Data scientists A. Saboni and A. Blangero explained that the more complex an algorithm becomes, the harder it gets to understand it.¹⁷⁶ They called it the performance/interpretability trade-off. If advertising is “normal business” for platforms, political advertisement is simply a socio-political use of social media.¹⁷⁷ Fingers have been pointed towards platforms, however they are not the only ones functioning behind closed doors. Following Cambridge Analytica, the European data protection authorities stressed that protection should not simply rely on the goodwill of the platforms.¹⁷⁸ The European Parliament had already been pushing for more data protection at the time. In fact, the General Data Protection Regulation (see section 6.4.1.a) was adopted on 14 April 2016 and entered into force in May 2018, which collided with the outbreak of the Cambridge Analytica/Facebook case.¹⁷⁹ Cambridge Analytica served as a warning for the EU

¹⁷³ EUROPEAN DATA PROTECTION SUPERVISOR, Opinion 3/2018 on online manipulation and personal data of 19 March 2018, <https://edps.europa.eu/sites/default/files/publication/18-03-19_online_manipulation_en.pdf> last accessed 15 July 2021, p.12.

¹⁷⁴ Charter of Fundamental Rights of the European Union, art. 8§2.

¹⁷⁵ F. PASQUALE, “Chapter 5: Machine Judging Humans”, in *New Laws of Robotics: defending human expertise in the age of AI*, Cambridge, Massachusetts & London, The Belknap Press of Harvard University Press, 2020, pp.119-144.

¹⁷⁶ A. SABONI & A. BLANGERO, *Opening the algorithm's black box and understand its outputs*, 4th International Conference on Electrical and Information Technologies, IEEE, 2020, <<https://ieeexplore.ieee.org/abstract/document/9113174>> last accessed 25 July 2021.

¹⁷⁷ S. WOOLLEY, *op. cit.* (voy. note 82), p. 87.

¹⁷⁸ EUROPEAN PARLIAMENT, “Artificial intelligence, data protection and elections”, 2019, <[https://www.europarl.europa.eu/RegData/etudes/ATAG/2019/637952/EPRS_ATA\(2019\)637952_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2019/637952/EPRS_ATA(2019)637952_EN.pdf)> last accessed 14 June 2021.

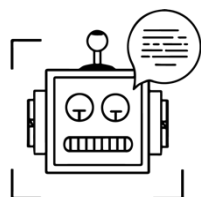
¹⁷⁹ I. NENADIC, *op.cit.* (voy. note 2), p.6.

and accelerated the call for regulators to “*make a joint effort to protect elections integrity*”.¹⁸⁰ As AI performance increases with time, the European Union will have to keep addressing how to maintain free and fair elections in an era where the rise of authoritarian powers co-exists with these types of new technologies.

c. Foreign intervention in national elections

Targeted political influence finally poses a threat to the principle of free and fair elections exempt from foreign interference. Again, the opacity of the technology and the advent of pervasive cross-border interconnectivity offer an opportunity for outsiders to interfere with free and fair democratic elections. Third-party interference in sovereign European elections “*constitutes a major challenge as it poses serious risks for European democratic societies and institutions, fundamental rights and freedoms, the rule of law, security, economic wellbeing and, ultimately, Europe’s sovereignty*”.¹⁸¹ This paper already described elections as an essential moment within democratic societies. In a way, the ballot is Rousseau’s “social contract” between the people and their representatives. Foreign intervention, whether through disinformation campaign, direct or indirect financial support or through pressure on certain actors breaches the values of the Union and the principle of sovereignty.

5.3. Political bots



Prior to election day, some voters also have received strange friend requests. Their newly acquired “friends” are surprisingly posting and reacting to a lot of political content online at rapid pace.

5.3.1. Political bots in action

The third AI-driven disinformation mechanism this paper addresses are political bots. The term “bot” is the contraction of the term “*software robot*”.¹⁸² Originally created to save time and energy, bots have existed for quite some time already.¹⁸³ It is an automated software which can interact with users or other systems. A chatbot has the specificity to engage conversationally with humans. The first

¹⁸⁰ EUROPEAN PARLIAMENT, *op. cit.* (voy. note 178), p.2.

¹⁸¹ Resolution (EU) 2019/2810 of the European Union of 10 October 2019 on foreign electoral interference and disinformation in national and European democratic processes, Recital C.

¹⁸² E. FERRARA et al., “The Rise of Social Bots”, *Commun ACM*, 59(7), 2016, <<http://dx.doi.org/10.1145/2818717>> last accessed 24 July 2021, p. 1.

¹⁸³ P. N. HOWARD, S. WOOLLEY & R. CALO, “Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration”, *Journal of Information Technology & Politics*, 15(2), 2018, pp.81-93, DOI: 10.1080/19331681.2018.1448735, p.82.

chatbot Eliza (a “psychotherapist”) was created by J. Weizenbaum in 1966.¹⁸⁴ The definition of a bot is ambiguous as it embraces a wide range of forms.¹⁸⁵ It can be run by a human, an algorithm or a combination of both. A distinction should be drawn between “dumb bots” and “smart bots”. Dumb bots are created for one specific task. They are rule-driven and will for example just share, like and comment certain content on social media. As opposed to dumb bots, smart bots are automated conversational agents using machine learning and deep learning to interact with people online. The AI programs are coded to imitate human interaction. Through conversations, the bot learns and improves its “reaction” towards becoming more and more humanlike. The software will consider human languages, analyse sentiments and make predictions based on statistics. The algorithm can optimise its interaction, learn from data and adapt its tactics. It is basically evolving towards passing the Turing Test.

5.3.2. The technology behind the humanlike nature of bots

Like deepfakes, the recent development of AI in bot creation raises a red flag. As S. Woolley wrote: “*conversational machine learning bots have existed, relatively speaking, for a long time. But until recently they required too many resources - not just money but also serious time and intellectual labour – to deploy in numbers*”.¹⁸⁶ Bots use deep learning technology which can analyse many texts rapidly. They can process natural language and assimilate a human communication style. Dumb bots and smart bots’ technology differs. Dumb bots use sequences of questions to fill in the linear decision-tree and conduct the one specific task it was coded for, while smart bots work in a web. To tend towards more humanlike bots, the entrepreneur Shane Mac introduced the Random-Access Navigation (or RAN).¹⁸⁷ As opposed to the rigidity of a decision-tree, the RAN technique takes into account more parameters (context, intent, etc.) and allows for a change of mind during the conversation with the bot (*web-like decision-making*), just as in any real human interaction. The technological differences between dumb bots and smart bots can be summarized as presented in the table below (see Figure 4 below). This paper will discuss one type of smart bot which is coded to pursue a political agenda: the political bot.

¹⁸⁴ R. KHAN & A. DAS, *Build Better Chatbots: A complete Guide Getting Started with Chatbots*, 1st ed., Apress, 2018, <<https://doi-org.proxy.bib.ucl.ac.be:2443/10.1007/978-1-4842-3111-1>> last accessed 25 July 2021, pp. 1-11.

¹⁸⁵ O. BOICHAK et. al., “Not the Bots You Are Looking For: Patterns and Effects of Orchestrated Interventions in the US and German Elections”, *International Journal of Communication*, 15, 2021, pp. 814-839, <<https://ijoc.org/index.php/ijoc/article/view/14866>> last accessed 17 August 2021, p.815.

¹⁸⁶ S. WOOLLEY, *op. cit.* (voy. note 82), p.87.

¹⁸⁷ S. MAC, “There are dozen ways to order a coffee. Why do dumb bots only allow one?”, Medium, 2017, <<https://medium.com/assist/theres-a-dozen-ways-to-order-a-coffee-why-do-dumb-bots-only-allow-one-27230542636d>> last accessed 24 July 2021.

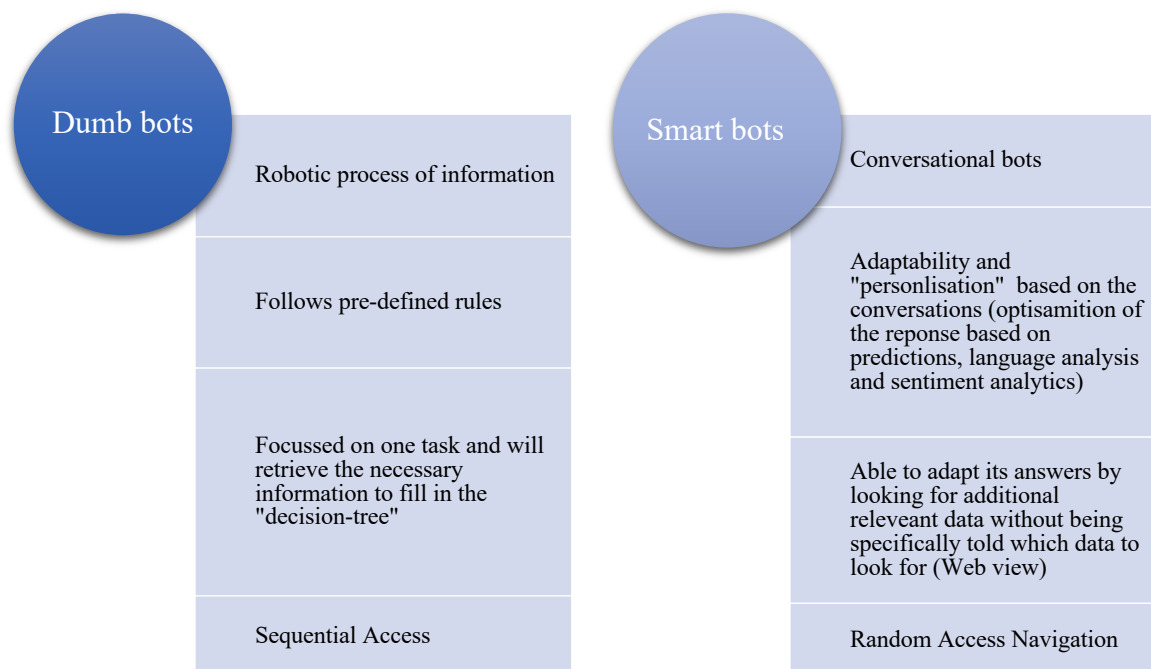


Figure 4 – Technological comparison of dumb bots and smart bots

5.3.3. Political bots and elections manipulation

The risk with political bots is that they can broadcast on a large scale false information as reflecting the opinion of citizens. The difference with targeted influence campaigns where the origin of the fake was unclear is that the message is directly connected to a “persona” in this case. The bot is “someone” relatable. The bot “persona” is getting more complex with fake background information, pictures and a language adapted to the intended audience. All these factors can help turn a bot into a weapon of democratic disruption.¹⁸⁸ AI and natural language processing make it harder to distinguish a bot from a real person. According to the latest Bad Bots report, malicious bots now constitute a quarter of the traffic on the Internet (15,2% for good bots’ traffic and 59,2% for human traffic).¹⁸⁹ Powered by a deep learning technology, bots can examine data and learn from it. They can determine with precision which information is interesting without any human intervention to prescribe which type on data it should focus on. When bots spread disinformation on upcoming elections under the identity of a fake and non-existent voter, it presents a risk of (a) disseminating misinformation online (b) taking the space and forum of real voters and (c) giving the impression that more people believe in certain ideas than is really the case.

¹⁸⁸ S. WOOLLEY, *op. cit.* (voy. note 82), p.24.

¹⁸⁹ IMPERVA THREAT RESEARCH LAB, “8th Annual Bad Bots Report”, <<https://www.imperva.com/resources/resource-library/reports/bad-bot-report/>> last accessed 26 July 2021.

a. Large scale propaganda

Bots, especially as a network of millions of accounts, can create a mass of fake followers to boost the apparent popularity of a candidate, start arguments online on polarising subjects, share propaganda under specific political hashtags, discuss with and influence voters, etc.¹⁹⁰ Political bots can also be used to manipulate information, amplify hate speech, collect data, troll or support a candidate.¹⁹¹ Again, bots are not new but the novelty comes from their evolution. As the Oxford Internet Institute wrote: “*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*”.¹⁹² The bots can now reach a larger audience while interfering with the democratic process. Bots (even dumb bots) can have an amplifying effect. This phenomenon is the consequence of the functioning of social media. We have seen that when a piece of content generates significant engagement (section 3.2.1), it is shared on more social media feeds. Bots have thus the capacity to enhance visibility by boosting engagement. Through bot reactivity, a political candidate or an interested third party have the capacity of creating a fictitious viral reaction to a given content by multiplying the bot’s engagement.

b. Altering the discussion equilibrium

At the same time and especially with the natural language processing, the line between human and robot is getting thinner. The bot “persona” is misleading, and it can be confused with a real person. The problem is that it typically amplifies voices already existing in the democratic debate instead of empowering voices struggling to be heard (women, minorities, etc.). To debate about a political topic, the opinion of the people is important. It should leave room to include the voices of every potential voter. Political bots take up the digital space. Abokhodair et al. mentioned that bots can either (re-)tweet certain opinions or disrupt communication by sharing under a hashtag unrelated content.¹⁹³ Aside from boosting a viewpoint, an automated army of bots can also harass the opposition or minority voices in an attempt to disrupt political discussions.¹⁹⁴ After the elections, the bot accounts are either deleted, going private or they are completely switching names and discussion topics.¹⁹⁵

¹⁹⁰ S. WOOLLEY, *op. cit.* (voy. note 82), p.24.

¹⁹¹ S. BRADSHAW & P. N. HOWARD, “The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation”, Computational Propaganda Research Project, Oxford Internet Institute, 2019, <<https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>> last accessed 20 July 2021, p.1.

¹⁹² S. BRADSHAW & P. N. HOWARD, *op. cit.* (voy. note 191), p.11.

¹⁹³ ABOKHODAIR N. et al., *Dissecting a social botnet: growth, content and influence in twitter*, Procedure of 18th ACM CSCW, 2015, <<https://arxiv.org/abs/1604.03627>> last accessed 3 July 2021, pp 839–851.

¹⁹⁴ S. WOOLLEY, *op. cit.* (voy. note 82), p.24.

¹⁹⁵ O. BOICHAK et al., *op. cit.* (voy. note 185), p.822.

Just like they spread and took space, they then shrink. Their tracks are covered and the digital evidence buried.

c. Generating a new narrative online

Bots can also give the impression that a politician has a bigger and more diverse support base than is real. They can pretend to represent minorities allegedly supporting and agreeing to hateful and racist ideas. The creation of these persona is part of a political strategy. Shao et al. established that bots tend to put forward low-credibility content.¹⁹⁶ But social media also plays a role. As explained (section 3.2.2), information cascade tends to vindicate the majority. Since we are now flooded with information, if the apparent majority believes something, the easiest thing to do is to go along because “it must be true”.¹⁹⁷ The actors behind political bot networks create a narrative and disseminate it into the social media bloodstream. Then, they let the bots discuss, share and influence online. Social bots were for example used to “guide” voters in US political campaigns.¹⁹⁸ In Europe, there was evidence of political bots use during the 2017 German elections and the 2016 UK referendum.¹⁹⁹

5.3.4. The impact of political bots

Just as for targeted influence campaigns, Brexit and the US elections have been subjected to political bots. F. Brachten et al. emphasised that these events have in common that pre-electoral polls predicted a totally different result.²⁰⁰ If it has been demonstrated that an important part of both campaigns occurred online, it is once again difficult to evaluate the impact of the bots' involvement. Political bots do amplify existing voices, but what is their real impact on election results? The quantitative empirical analysis of C. Shao et al. on low-credibility content pinpointed two bots' manipulative strategies: early amplification of existing content and targeting influential users through replies and mentions. The findings of the study concludes that bots are part of the mechanisms

¹⁹⁶ C. SHAO et al., “The spread of low-credibility content by social bots”, *Nature Communication*, 9, 2018, <<https://doi.org/10.1038/s41467-018-06930-7>> last accessed 28 July 2021.

¹⁹⁷ D. EASLEY and J. KLEINBERG, *Networks, Crows, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, 2010, <<https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf>> last accessed 18 July 2021, p.484.

¹⁹⁸ J. BAYER et al., *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States*, Study Requested by the LIBE committee, <[https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU\(2019\)608864](https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2019)608864)> last accessed 20 July 2021, p.119.

¹⁹⁹ O. BOICHAK et al., *op. cit.* (voy. note 185), p. 816.

²⁰⁰ F. BRACHTEN et al., *Strategies and Influence of Social Bots in a 2017 German state election: a case study on Twitter*, Australasian Conference on Information Systems, 2017, Hobart, <https://www.researchgate.net/publication/320564398_Strategies_and_Influence_of_Social_Bots_in_a_2017_German_state_election_-_A_case_study_on_Twitter> last accessed 20 July 2021, pp.2-3.

spreading misinformation (next to algorithms and cognitive limitations like echo chambers). Similarly, Vosoughi et al. argued that bots cannot explain alone the rise of online misinformation, but that they participate to the current misrepresentation era.²⁰¹ Additionally, S. Woolley in his study about social bots' interference in politics proved that the more a country is democratic, the more bots will only or mostly be used for follower padding.²⁰² Authoritarian countries will resort to bots for other purposes such as pro-government messages and the demobilisation of the opposition. Thus, it is a vicious circle: not protecting democracy is leaving room for more impactful political bots and political bots can further endanger democracy. The current rise of authoritarian regimes within the EU (section 2.3.2) is paving the way for new forms of computational propaganda.

In contrast, a study conducted by a consortium of researchers from the U.S. and Australia established that bots are not effective to influence individuals whose opinions are completely divergent.²⁰³ According to O. Boichak et al., there are three key dimensions of information diffusion: speed, range and scale. As much as bots create an amplifying (*scale*) and viral effect (*speed*), the empirical data of the study showed a failure to reach a new audience (*range*).²⁰⁴ Hence, researchers do not necessarily agree on the extent to which political bots may influence elections and harm democracy. But they do agree that even if bots are not necessarily an effective stand-alone weapon, they are most definitely part of the disinformation arsenal.

5.3.5. Political bots and fundamental rights

a. Right to non-discrimination

As a disinformation tool and as a type of computational propaganda, the political participation of bots in an electoral process triggers similar fundamental rights to those already discussed. This section will thus rely on what this paper has previously covered. Similarly to discriminatory and harassing deepfakes (see section 5.1.4.b), political bots can be created with the aim of silencing dissenting or unwelcome voices (e.g. minorities, women, a political opponent). The use of bots in political campaigning to attack and harass minorities is breaching the right to non-discrimination of article 21 EU Charter and article 14 ECHR. If bots are not illegitimate *per se*, political bots should not interfere with the equal participation of the people to the democratic debate. This breach would

²⁰¹ S. VOSOUGHI, D. ROY & S. ARAL, *op.cit.* (voy. note 61), pp.1146-1151.

²⁰² S. C. WOOLLEY, "Automating power: social bot interference in global politics", *First Monday*, 21(4), 2016, <<https://doi.org/10.5210/fm.v21i4.6161>> last accessed 19 July 2021.

²⁰³ O. BOICHAK et al., *op. cit.* (voy. note 185), pp.817-818.

²⁰⁴ O. BOICHAK et al., *op. cit.* (voy. note 185), pp.817-818.

entail the use of the bot technology in a discriminatory fashion with the intention of deceiving certain groups of citizens and influencing their political behaviour.²⁰⁵

b. Fair and transparent electoral process

Pluralism is a characteristic of a healthy democracy, meaning that citizens must be able to access a variety of information to form their political opinion.²⁰⁶ But what political bots working as one unit offer is a weapon to help conduct an imbalanced and unfair electoral campaign. It provides a mass power able to saturate communications online. It amplifies one standpoint and shadows the visibility of other political opinions. The improvement of the technology makes it harder to detect when an opinion is defended by a bot versus when the argument stems from a peer. Whether at national or European level, there are rules on transparency and fair political campaigning (e.g. a candidate may only spend a given amount of money for advertising). But as political bots become harder to distinguish and harder to trace back to their actual origin, they also become harder to subject to transparency rules.

c. Foreign intervention in national elections

Finally, political bots challenge the right to conduct free and fair elections away from foreign intervention. It is an accepted fact that Russia interfered in 2016 U.S. elections.²⁰⁷ Similarly, in 2017, a disinformation campaign was launched by Sputnik, RT France (*previously “Russia Today”*) and a network of bots on Emmanuel Macron’s presidential campaign.²⁰⁸ The attack was exposed before it could reach its purpose. This illustrates however the existence of the threat and the need for and usefulness of detecting devices. Political bots constitute a weapon of interference in elections in which foreign states may have personal interest in the win of a particular candidate. This circles back to the fact that as the detection of smart bots is getting more complex, exposing (foreign) interference is also getting more challenging.

²⁰⁵ I. NENADIC, *op.cit.* (voy. note 2), p.5.

²⁰⁶ I. NENADIC, *op.cit.* (voy. note 2), pp.4-5.

²⁰⁷ M. LATONERO, *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, Data&Science, 2018, <<https://apo.org.au/sites/default/files/resource-files/2018-10/apo-nid196716.pdf>> last accessed 14 July 2021, p.12.

²⁰⁸ European Parliament, “Foreign influence operations in the EU”, Briefing, <[https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI\(2018\)625123_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI(2018)625123_EN.pdf)> last accessed 19 July 2021, p.7.

5.4. Section conclusion: the growth of emotions-centred political strategy

These three AI-driven technological phenomena are rooted in disinformation. Whilst the public now knows about misinformation (more often under the well-known and controversial term “fake news”), it does not prevent them from being exposed to and influenced by it. The difference with “traditional disinformation” is that deepfake and computational propaganda (political bots and targeted influence campaigns) combine AI technology, misinformation techniques and the broadcast power of social media. These AI disinformation tools also direct the attention towards emotional reactions and away from political facts. What the Cambridge Analytica scandal demonstrated is that winning an election has become predicated upon winning the emotional argument behind a collective issue, irrespective of whether or not the argument is rooted in facts. It is about speaking to and feeding into existing emotions and fears. The risk of demagoguery already highlighted by Plato (428-348 BC) is materialised even more when boosted by new technologies and social media. The attention should return to the real issues facing democratic societies. As N. Nyabola urges, the current climate needs to bring human and truth back at the centre of the electoral process.²⁰⁹ But how do we achieve this ambitious mission? In 2020, the European Union issued its European Democracy Action Plan to “*empower Europeans and to strengthen the resilience of our democracies*”.²¹⁰ It is a testimony of its willingness to ensure “*healthy and thriving*” democracies across the EU. According to Professor D. Citron, the solution requires diverse actions. She advocates for a combination of law, market action, and societal resilience.²¹¹ While these three areas are equally important to protect free and fair elections, the focus of this paper will remain on the European legal response, specifically the solutions offered by the European Commission’s recent AI Regulation proposal.

6. Artificial Intelligence Act: perspectives for AI-related election manipulation

This section will analyse the above-mentioned issues regarding free and fair elections from the angle of the European Commission’s 2021 AI Regulation proposal. The aim is not to summarise all key elements of the proposal but rather to provide a context and see how the draft is discussing potential solutions with regards to AI-related election manipulation.

²⁰⁹ SEF: DEVELOPMENT AND PEANCE FUNDATION, “We need to bring the human back into the digital conversation”, Interview with Nanjala Nyabola on social media and its role in Kenya sef: insight 2, 2019, <<https://www.sef-bonn.org/en/events/dresden-forum-for-international-politics/2019.html>> last accessed 5 July 2021.

²¹⁰ EUROPEAN COMMISSION, “European democracy Action Plan: Making democracies stronger”, Factsheets, 2020, <https://ec.europa.eu/info/sites/default/files/edap_factsheet8.pdf> last accessed 3 July 2021.

²¹¹ HOUSE INTELLIGENCE COMMITTEE, intervention of D. CITRON at the National Security Challenges of Artificial Intelligence, Manipulated Media, and Deepfakes, 116th Congress, 2019, <<https://www.congress.gov/event/116th-congress/house-event/109620>> last accessed 3 June 2021.

6.1. Context of the AI Act Proposal

Aspiring to be a major initiator of policy and legislative initiatives for the digital sector, the European Union has put forward several instruments in the field. In February 2020, the European Commission published its White Paper on Artificial Intelligence,²¹² paving the way for a future regulatory framework. On 21 April 2021, the European Commission suggested the first-ever legal framework on AI: the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts.²¹³ The Commission opted for a regulation with the desire to see the new rules uniformly applied throughout the Union.²¹⁴ Indeed, a regulation is binding and directly applicable in all Member States.

The proposal was presented in parallel with other draft policy initiatives including the Digital Services Act, the Digital Markets Act and the Data Governance Act.²¹⁵ It is part of a broader programme that encompasses an educational and a training feature in key digital technologies and artificial intelligence.²¹⁶ Hence, AI is now well-established on the EU's political agenda. AI was in fact part of President von der Leyen's political commitments before she officially took office.²¹⁷ As a proposal, the AI Act still needs to go through the ordinary legislative process (art. 294 TFEU) before entering into force as amended. At the time of writing, the draft is going through its first reading. It has already been discussed by the Council and its preparatory bodies on several occasions but has yet to be debated in the European Parliament.²¹⁸

²¹² EUROPEAN COMMISSION, White paper on Artificial Intelligence: a European approach to excellence and trust of 19 February 2020, COM (2020) 65 final.

²¹³ Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) amending certain union legislative acts of 21 April 2021, COM (2021) 206 final, hereinafter: "Artificial Intelligence Act of 21 April 2021".

²¹⁴ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.7.

²¹⁵ M. VEALE AND F. Z. BORGESIU, "Demystifying the Draft EU Artificial Intelligence Act", *Computer Law Review International*, 22(4), 2021, <<https://arxiv.org/abs/2107.03721>> last accessed 25 July 2021, p.3.

²¹⁶ EUROPEAN COMMISSION, The performance framework for the EU budget, under the 2021-2027 multiannual financial framework, volume II, Annex, 2021, <<https://op.europa.eu/en/publication-detail/-/publication/1d4ad9c8-d23b-11eb-ac72-01aa75ed71a1/language-en/format-PDF/source-search>> last accessed 14 July 2021, p.26.

²¹⁷ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.1.

²¹⁸ Procedure 2021/0106/COD COM (2021) 206 for the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) amending certain union legislative acts, <https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX:52021PC0206#2021-04-22_ADP_byCOM> last accessed 14 July 2021.

6.2. Objectives of the AI Act: the balance between innovation and safeguards

The European cross-sectoral strategy regarding AI intends to both encourage AI innovation and provide a legal framework for it.²¹⁹ If this master thesis has for the most part highlighted the risks posed by AI, there are also positive outtakes. Advantages range from economic growth to innovation and societal benefits.²²⁰ But as recital 15 of the proposal for Artificial Intelligence Act (hereinafter: “AI Act” or “draft AI Act”) mentions: “*aside from the many beneficial uses of AI, that technology can also be misused and provide novel and powerful tools for manipulative, exploitative and social control practices*”.²²¹ The proposal for the AI Act provides a harmonised definition of AI,²²² a risk-based classification of certain AI systems and the prohibition of certain AI-related practices.

Therefore, the Artificial Intelligence Act proposed by the European Commission aims at:

- (i) *ensuring that systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;*
- (ii) *Ensuring legal certainty to facilitate investment and innovation in AI;*
- (iii) *Enhancing governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;*
- (iv) *Facilitating the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.*²²³

The objective is to reach a “*balanced and proportionate regulation*”.²²⁴ It means having a strong response to high-risk AI technologies while still promoting the development of AI applications across the EU. The challenge is tremendous. The European Union has only 6 of the top 100 AI start-ups of the world.²²⁵ Regulation should not scare away businesses and investors. The internal market must stay appealing despite the prospect of new safeguards being set in place. During the drafting process, the Commission was in fact opened to hearing the market’s perspective. It conducted a stakeholders’

²¹⁹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art.3.

²²⁰ EUROPEAN PARLIAMENT, “Legislative Train 06.2021, A Europe fit for the digital age: proposal for a regulation on a European Approach for Artificial Intelligence”, <<https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>> last accessed 11 July 2021.

²²¹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.21.

²²² Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art.3.

²²³ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.3.

²²⁴ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.3.

²²⁵ CBINSIGHTS, “AI 100: The Artificial Intelligence Startups Redefining Industries”, 2021, <<https://www.cbinsights.com/research/report/artificial-intelligence-top-startups/>> last accessed 10 July 2021.

consultation where written feedback was provided via an online questionnaire.²²⁶ Despite this consultative approach, criticism was levelled at the eventual proposal. Dr. B. Mueller expressed the doubt that such regulation could achieve the intended aims. He regrets the “[*micro-managing*] of the use of AI across a vast range of applications rather than focusing on actual threats like (...) *disinformation*”. According to Mueller, the draft is a step back from the ambition to turn the EU into a worldwide AI leader as it would discourage AI start-ups to innovate in this sector within the Union.²²⁷

Undeniably, striking a balance between regulatory robustness and market flexibility is challenging. The question as to whether or not the EU’s “ethics-first” approach is raising a self-inflicted obstacle in the technological race with the U.S. and China is still ongoing. But previous instruments have demonstrated that the bold policy-making ambitions of the EU can lead to setting new international standards. The General Data Protection Regulation has influenced the improvement of global data protection rules. Similarly, the AI Act could have a wide-reaching regulatory impact on how to deal with AI systems in the future.

6.3. The risk-based approach

6.3.1. Temporal, territorial and material scopes

The official date of the entry into force of the instrument is not yet specified by Article 85 of the proposal since it still requires a co-decision of the Council and the Parliament. According to the EU financial statements, the instrument is part of the 2022 budget.²²⁸ The material scope of the regulation covers the “*placing on the market, the putting into services and the use of AI systems*”.²²⁹ The proposal as it stands would be applicable to any stakeholder (whether public or private), including financial services, irrespective of whether these providers of AI systems are established within the EU or in third countries.²³⁰

²²⁶ EUROPEAN COMMISSION, White Paper on Artificial Intelligence: a European approach to excellence and trust, consultation, 2020, <<https://digital-strategy.ec.europa.eu/en/consultations/white-paper-artificial-intelligence-european-approach-excellence-and-trust>> last accessed 14 July 2021.

²²⁷ B. MUELLER, *The Artificial Intelligence Act is a Threat to Europe’s Digital Economy and Will Hamstring the EU’s Technology Sector in the Global Marketplace*, Centre for Data Innovation, Press Release, 2021, <<https://datainnovation.org/2021/04/the-artificial-intelligence-act-is-a-threat-to-europes-digital-economy-and-will-hamstring-the-eus-technology-sector-in-the-global-marketplace/>> last accessed 12 July 2021.

²²⁸ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), pp.91-103.

²²⁹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.12.

²³⁰ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 2.

6.3.2. First harmonised definition of AI

The scope of the instrument refers to “AI systems”. In fact, the regulation offers the first harmonised definition of AI, as requested by the respondents to the public consultation.²³¹ Article 3(1) AI Act says AI systems are “*software that are developed with one or more of the techniques and approaches listed in the Annex I, and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with*”. P. Glauner criticised the broad definition as it would encompass almost any software.²³² So far, the list of Annex I covers techniques involving machine learning, statistical approach as well as logic and knowledge-based approach.²³³ This open-ended definition is meant to be future-proof.²³⁴ AI being a fast-changing area, the European Commission took into account its evolutionary nature. The risk of a closed-end definition would be obsolescence. As a result, Article 4 AI Act allows the Commission to amend and update Annex I. The above-evaluated AI tools (see section 5) would already all be covered in the proposed definition of AI systems. Deepfakes and computational propaganda are software-generated contents involving machine learning influencing the environments they interact with.

6.3.1. The pyramid of risks

Introducing a risk-based approach, the Commission has categorised risks into four categories: unacceptable risks, high risks, limited risks and minimal risks. The AI election manipulation tools being within the scope of application of the draft, the next question is: under which category will they fall? This paper will not analyse minimal risk AI systems but will look at how the tools might be labelled under unacceptable, high and limited risks.

²³¹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.8.

²³² P. GLAUNER, “An Assessment of the AI Regulation Proposed by the European Commission”, preprint to appear in 2022 in *The Future Circle of Healthcare: AI, 3D, Printing, Longevity, Ethics and Uncertainty Mitigation* (eds. S. EHSANI et al.), 2022 <<https://arxiv.org/pdf/2105.15133.pdf>> last accessed 10 July 2021, pp.3-4.

²³³ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), Annex I.

²³⁴ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.3.

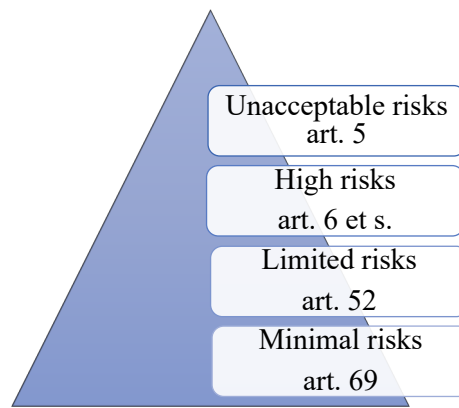


Figure 5 - The pyramid of risks of the AI Act draft

a. Unacceptable risks: prohibited AI practices

Some AI systems would be prohibited from being put on the market, put into service or used altogether if they caused physical or psychological harm. Article 5 exhaustively lists the following: techniques able to distort someone's consciousness subliminally; techniques exploiting the vulnerabilities of a specific group due to their age, physical or mental disability; techniques of social scoring leading to detrimental or unfavourable treatment and techniques of real-time remote biometric identification systems in publicly accessible systems for the purpose of law enforcement.²³⁵

The prohibition of certain unacceptable AI practices aims at protecting the Union's values and safeguarding fundamental rights.²³⁶ However, if section 5 of this paper has demonstrated that deepfakes, targeted political campaigning and political bots could infringe fundamental rights, the evidence of psychological harm resulting from targeted political campaigning and political bots is harder to demonstrate. For example, proving that someone's consciousness has been distorted subliminally would be hardly possible to prove. It would result in discussions about the real impact of AI on elections results. The Cambridge Analytica case showed that even experts disagree on how much influence the company actually had. Without a method to quantify the impact, similar inconclusive deliberations would arise.

As for deepfakes, if a victim can easily argue that he or she suffered psychological harm, citizens whose ballot would have been influenced by a fake would have a hard time demonstrating suffering psychological harm as a direct consequence. The risk-based approach does not include a risk to

²³⁵ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 5. *Techniques of real-time remote biometric identification systems in publicly accessible systems for the purpose of law enforcement accepts certain exceptions. See article 5(d) AI Act.*

²³⁶ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.12.

society or democracy. Unfortunately, the draft lacks a “collective harm” criterion that would enable a broader application (see section 6.4.2).²³⁷ Therefore, even if specific circumstances might fall under the unacceptable risks category (e.g. AI tool of election manipulation specifically targeting and exploiting the vulnerabilities of mentally disabled persons), most instances will remain complex to argue and prove.

b. High-risk and mandatory requirements

The second option would be to argue that (political) deepfakes, targeted campaigning and political bots are part of what Article 6 refers to as high-risk AI systems.²³⁸ These systems are those posing “significant risks” to health and safety or fundamental rights.²³⁹ High-risk AI system would operate lawfully under the condition that they fulfil defined requirements prior their access to the market (risk assessment and mitigation systems;²⁴⁰ high quality of datasets feeding the system to minimize risks and discriminatory outcomes;²⁴¹ record-keeping of activity to ensure traceability of results;²⁴² detailed documentation providing all information necessary on the system and its purpose for authorities to assess its compliance;²⁴³ clear and adequate information to the user;²⁴⁴ appropriate human oversight measures²⁴⁵ and high level of robustness, security and accuracy²⁴⁶). The category of high-risk AI systems and its underlying obligations constitutes the majority of the provisions of the proposal. The category encompasses AI systems intended to be used as a product (or safety component thereof) covered by EU harmonisation legislation and required to undergo a third-party conformity assessment.

In addition, AI systems referred to in Annex III are also considered as high-risk. The list of Annex III mentions eight fixed categories of applications ((i) biometric identification and categorisation, (ii) management and operation of critical infrastructure, (iii) education and vocational training, (iv) employment, worker management and access to self-employment, (v) access to and enjoyment of essential private services and public services, (vi) law enforcement, (vii) migration and border control

²³⁷ M. VEALE AND F. Z. BORGESIU, “Demystifying the Draft EU Artificial Intelligence Act”, *Computer Law Review International*, 22(4), 2021, <<https://arxiv.org/abs/2107.03721>> last accessed 25 July 2021, pp.4-5.

²³⁸ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), Title III.

²³⁹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.3.

²⁴⁰ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 9.

²⁴¹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 10.

²⁴² Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 12.

²⁴³ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 11.

²⁴⁴ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 13.

²⁴⁵ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 14.

²⁴⁶ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 15.

management, (viii) administration of justice and democratic processes), none of which are fit for addressing AI election manipulation tools. Indeed, category viii has only one sub-category covering “*AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to concrete set of facts*”.²⁴⁷

As it stands, the high-risk AI systems category does not appear to embrace deepfakes and computation propaganda. However, Article 7 AI Act empowers the Commission to amend Annex III in order to add new sub-categories to any of the eight categories under the condition that it poses a risk of harm to health and safety or a risk of adverse impact on fundamental rights.²⁴⁸ The Commission will have to assess the severity and probability of occurrence taking into account the purpose of the system, the extent to which it has been used, the extent of the harm and the extent to which Union legislation provides for effective measure of redress or prevention.²⁴⁹

If introduced in the current form, the regulation would not address deepfakes, targeted political campaigning and political bots under the definition of high-risk systems. While the Commission has the power to update the list of Annex III, it would have to prove that deepfakes and computation propaganda harm fundamental rights, have “*the potential to affect a plurality of persons*” and do not offer the “*possibility for users to opt-out from the adverse outcome*”. The amendment would also have to fit into one of the pre-existing categories of Annex III. No completely new areas will be admitted.²⁵⁰

c. Limited risks and the transparency requirement

The final opportunity of the draft to protect free and fair elections from AI elections manipulation is the category of limited risks. Title IV of the regulation establishes an obligation of transparency for AI systems presenting specific risks of manipulation. Article 52 sets out an obligation of information for AI systems (i) interacting with natural persons (*relevant to bots*) (ii) used to detect emotions or determine association with (social) categories based on biometric data (*relevant to targeted influence campaign*) (iii) generate or manipulate content (*relevant to deepfake*).

²⁴⁷ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), Annex III 8(a).

²⁴⁸ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 7.

²⁴⁹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 7(2).

²⁵⁰ M. VEALE AND F. Z. BORGESIU, *op. cit.* (voy. note 237), p.10.

Article 52(1): Disclosure of bots

The European ambition to improve AI disclosure is not new.²⁵¹ The European Commission's 2018 Code of Practice on Disinformation was already aiming at “*establish clear marking systems and rules for bots and ensure their activities cannot be confused with human interactions*”.²⁵² In this instance, Article 52(1) AI Act specifically refers to chatbots “providers” who will have an obligation of disclosure.²⁵³ For example, a bot will have to present a label indicating that the interaction is not conducted by a human being.²⁵⁴ However, this obligation has limitations. No disclosure is required if the AI-nature of the system is obvious or if the content is justified by legitimate purposes. Moreover, the obligation does not include an obligation for users or platforms and only concerns providers. Providers are “*natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge*”.²⁵⁵ By contrast, users are using an AI system under their authority.

M. Veale and F. Z. Borgesius regret the absence of broader transparency obligations.²⁵⁶ They consider that “providers” and “users” are often intertwined in practice. The proposal assumes that providers create AI systems and resell them as is. But as the technology evolves and becomes easier to use, the distinction is getting less meaningful. Veale and Borgesius argue that a hybrid solution would be preferable, and they therefore call for the suppression of the distinction between “user” and “provider” in Article 52. Moreover, in case a bot were to share false information in a bid to manipulate an election, it is not clear whether knowing the AI-nature of the bot would materially change the way in which individuals interact with the content in question. Bot disclosure does not imply that the content the bot shares will be removed. The (dis)information could still be occur.

Article 52(2): Emotion recognition and biometric categorisation disclosure

Article 52(2) AI Act states that “users” of an emotion recognition system and a biometric categorisation system have a duty to disclose the use of such AI system. Biometric data are data

²⁵¹ M. VEALE & F. Z. BORGESIOUS, *op. cit.* (voy. note 237), p.10.

²⁵² EUROPEAN COMMISSION, Tackling Online Disinformation: A European Approach, COM/2018/236 Final, 2018, <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>> last accessed 10 July 2021, p.8.

²⁵³ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), p.3.

²⁵⁴ S. F. SCHWEMER et al., *Legal AI Systems in the EU's proposed Artificial Intelligence Act*, Legal AIIA, 2021, <<http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>> last accessed 16 July 2021, p.3.

²⁵⁵ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 3(2).

²⁵⁶ M. VEALE & F. Z. BORGESIOUS, *op. cit.* (voy. note 237), 17-18.

registering human physical, chemical or behavioural characteristics.²⁵⁷ For example, a biometric data would be the fingerprint used to unlock a smartphone or authenticate an online payment. The user of the AI system will have to inform the individuals being categorised that the processing is performed via AI. The provision includes exceptions for “*AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences*”.²⁵⁸ As Article 52(2) mostly concerns personal data, the question is however how it relates to data protection as covered by the GDPR. The proposal will have to be integrated into and articulated with existing legislation (see section 6.4.1).

Article 52 is not the only article addressing biometric data. Under Article 5, the proposal establishes that types of biometric technology uses are purely prohibited.²⁵⁹ But it concerns only “real-time” remote biometric identification systems used for law enforcement. The limited ban of Article 5 does not concern other types of biometric systems. The obligation of Article 52(2) however is far less burdensome as it requires simple disclosure. Even though the biometric systems excluded from the scope of Article 5 would be covered by other EU instruments (typically the GDPR), the proposal misses an opportunity to address these other problematic uses of biometric data. Biometric categorisation combined with other data (e.g. political orientation) has proven to be useful to map a population into influenceable sub-groups. According to AI expert D. Leufer, the treatment of biometric categorisation is very problematic in the proposal.²⁶⁰ Knowing the potentiality of discriminatory categorisation or behavioural predictions, biometric categorisation should not be limited to a minimal transparency requirement. The proposal does not come close to addressing the problem upfront or to offering a viable solution to curb targeted political campaigning. It will require to articulate the proposal with the already-existing legislation.

Article 52(3): Deepfake disclosure

Article 52(3) establishes an obligation towards “users” to disclose when a content has been artificially altered.²⁶¹ There are again exceptions justifying an exemption, including systems authorised by law

²⁵⁷ I. NATGUNANATHAN et al., *Protection of Privacy in Biometric Data*, IEEE Access, 2016, <<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7420576>> last accessed 20 July 2021.

²⁵⁸ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 52(2).

²⁵⁹ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 5(1)(d).

²⁶⁰ S. CHANDER & E. JAKUBOWSKA, *EU’s AI law needs major changes to prevent discrimination and mass surveillance*, EDRI, 2021, <<https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/>> last accessed 1 August 2021.

²⁶¹ EUROPEAN PARLIAMENT, “Parliamentary questions: answer E-001309/2021 given by Mr. Breton on behalf of the European Commission”, 2021, <https://www.europarl.europa.eu/doceo/document/E-9-2021-001309-ASW_EN.html> last accessed 10 July 2021.

to detect, prevent, investigate and prosecute criminal offences or systems necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences.²⁶² A satirical deepfake could for example justify such an exemption from disclosure. Like for bot disclosure, M. Veale and F. Z. Borgesius questioned the distinction between “provider” and “user”. If the final text maintains the distinction, it will have to clarify when a provider becomes a user in order to determine what triggers the applicability of the provision.

The minimal transparency obligation raises questions as to whether this information is sufficient to prevent harm. Indeed, Article 52 only creates a right to be informed of the nature of the system but not about the nature of the content it shares. Knowing that it is a deepfake does not mean that the content will be disregarded. And as the content will still exist, so will the social and algorithmic phenomena (like the information cascades), the exposure to thought-confirming content or the fact that false information is often more appealing (section 3.2). So, is mere transparency (through communication) enough to tackle AI-based elections manipulation?

6.4. AI election manipulation: Assessment of the AI Act draft

After explaining the risk-based approach of the proposal and examining to what extent AI elections manipulation tools would be tackled, this section will assess the draft in the light of the protection of free and fair elections.

6.4.1. The challenge of articulation with other instruments

a. Data protection and the compliance with the GDPR

In its assessment paper, P. Glauner expressed the opinion that the proposal poses a risk of overregulation.²⁶³ Beside the risk of deterring innovation, overregulation may be critical for legal certainty. As the proposal regulates topics already covered by other instruments, interaction between those instruments should be duly taken into account in the drafting. One of these instruments is the GDPR.

Data whether personal or non-personal is often at the heart of any automated decision. The GDPR thus applies to any AI system resorting to personal data. As stated by the European Data Protection Board, any “*processing of personal data through an algorithm falls within the scope of the*

²⁶² European Charter of Fundamental Rights, art. 13.

²⁶³ P. GLAUNER, *op. cit.* (voy. note 232), p.3.

GDPR”.²⁶⁴ In 2018, the Commission issued a Guidance on the application of Union data protection law in the electoral context.²⁶⁵ In their joint opinion on the proposal, the European Data Protection Board and European Data Protection Supervisor (hereinafter: “EDPB” and “EDPS”) stressed that the draft AI Act has “*prominently important data protection implications*”.²⁶⁶ So, how to ensure the correct articulation between the AI Act and the GDPR?

The two instruments have many aspects in common. Firstly, following the governance model of the GDPR and its related EDPB and EDPS, the draft AI Act establishes a European AI board. The Board will be chaired by the Commission and oversee the application and implementation of the regulation. Secondly, the rules and their sanctions have an extended scope of application. They extend to providers and users including if established outside of the EU if their systems operate within the EU.²⁶⁷ But the draft AI Act opted for a risk-based approach rather than a rights-based approach. It does not create new rights. The protection of fundamental rights prescribed by the regulation should be done in consistency with the EU Charter of Fundamental Rights, Union legislation on data protection, the General Data Protection Regulation, non-discrimination and gender equality. It shall not encroach upon existing rights and obligations.

The EDPB and the EDPS joint opinion emphasised that a high-risk AI system under the AI Act will not necessarily be lawful (even though it might correctly respect its obligations under the regulation). It will need to comply with further EU data protection law. The articulation with the GDPR is thus very important. For example, the restriction of biometric AI systems of the draft AI Act only concerns AI systems used for law enforcement.²⁶⁸ Other types of AI systems using biometric data will need to be referred to the GDPR to be assessed. The EDPB and the EDPS also recommends that the concept of “risk to fundamental rights” be clearer and more aligned with the GDPR. The fact that the text

²⁶⁴ EDPB-EDPS, Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 2021, <https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf> last accessed 19 July 2021, p.8.

²⁶⁵ Commission guidance on the application of Union data protection law in the electoral context, a contribution from the European Commission to the Leaders’ meeting in Salzburg of 19-20 September 2018 COM/2018/638 final.

²⁶⁶ EDPB-EDPS, *op. cit.* (voy. note 264), p.2.

²⁶⁷ *The Draft AI Act establishes an obligation for member states to lay down rules on penalties. Article 71 provides that a ban breach will be punishable of a fine up to 30 000 000 EUR or 6% of the worldwide annual turnover. A breach of high risks obligations will be fined up to 20 000 000 EUR or 4% of the turnover. The supply of incorrect, incomplete or misleading information to notified bodies and national competent authorities in reply to a request shall be subject to administrative fines of up to 10 000 000 EUR or 2% of the worldwide turnover (whichever is higher).*

²⁶⁸ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art. 5(1)(d).

makes no reference to individuals affected by the AI systems is a “*blind spot in the proposal*”.²⁶⁹ They advocate for the creation of rights and remedies for victims of AI systems.

b. The Digital Services Act and illicit content online

The Digital Services Act (hereinafter: “DSA”) must also be mentioned.²⁷⁰ Creating a coherent digital strategy requires an articulation with the DSA. The DSA sets out “*rules for very large platforms to assess and mitigate the negative effect and systemic risks of their services to intentional manipulation*”.²⁷¹ The purpose of the DSA is to create a safer digital space by protecting the fundamental rights of digital services users. The instrument introduces rules against harmful content online. The DSA is applicable for digital service providers such as social media or online platforms in the EU or whose users are within the EU.²⁷² In the context of elections manipulation, the DSA would require flagging and removing illegal content. This includes for example AI-generated harmful deepfake or illegal adverts targeting individuals. In this regard, the draft AI Act specifies under article 2 that the AI Act should not affect the application of the DSA regarding the liability of intermediary services providers.²⁷³

The GDPR and DSA examples demonstrate that other EU instruments may apply to elections manipulation. A discussion on how these instruments may address the matter in combination would certainly be very interesting, but it is too far reaching given the boundaries of this paper. However, it is worth mentioning that legislators drafting the AI Act will have to carefully consider those other instruments. They will have to ensure that the interaction between the different texts does not compromise legal certainty (especially on a sensitive democratic topics).

6.4.2. The lack of a societal risk against discrimination and electoral disruption

The draft AI Act ambitions to be a pioneer in AI regulation. But despite the very broad definition of “AI systems”, the draft fails to address specifically certain elements that could have had a greater impact against the risk of elections manipulation. First, as the EDPB and EDPS joint opinion highlighted, the proposal leaves out the “*risk for groups of individuals or the society as a whole*”.²⁷⁴

²⁶⁹EDPB-EDPS, *op. cit.* (voy. note 264), p.8.

²⁷⁰ Proposal for a Regulation of the European Parliament and the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final.

²⁷¹ EUROPEAN PARLIAMENT, “Parliamentary question E-001309/2021: answer given by Mr. Breton on behalf of the European Commission,”, 2021, <https://www.europarl.europa.eu/doceo/document/E-9-2021-001309-ASW_EN.html> last accessed 25 July 2021.

²⁷² Proposal for a Regulation of the European Parliament and the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final, art. 1.

²⁷³ Artificial Intelligence Act of 21 April 2021, *op. cit.* (voy. note 213), art.2(5).

²⁷⁴ EDPB-EDPS, *op. cit.* (voy. note 264), pp.8-9.

This risk should also be assessed and mitigated, particularly bearing in mind the democratic impact of such harm. Researchers find the notion of “harm” of Article 5 (*prohibited AI practices*) to be vague and unfortunately limited to individual user harm.²⁷⁵ The addition of the notion of “collective harm” could for example include discrimination or electoral disruption, which would make more sense in the current EU political climate. A scandal like Cambridge Analytica has highlighted the existence of online vulnerabilities, and it can be argued that the proposal missed a step to be truly ground-breaking on the subject.

Second, P. Glauner regrets the absence of social media regulation “*even though this is precisely where there is a realistic danger of AI influencing our society*”.²⁷⁶ The draft was in fact toned down. So, if the definition of “AI systems” includes almost any software, the different risk categories overlook certain very harmful artificially altered content. This, combined with the absence of a notion of a “risk to society” gives the impression that AI election manipulation tools were not envisioned as the intended target by the Commission’s draft. The absence of societal risk in the proposal challenges even more the necessity to create a coherent articulation with other EU instruments (see section 6.4.1). If certain obligations with regards to bots, deepfakes or targeted campaigning will rise from the AI Act, simply complying with it might not be enough. The AI tools will have to be developed and used in accordance with the entire EU framework. Making the articulation of the framework as clear as possible from the get-go could avoid confusion and harmful results.

6.4.3. Checks and balances of fundamental rights

The eventual application of the draft AI Act will also need to balance the use of AI systems with innovation objectives and the safeguard of fundamental rights. The precautionary approach of the instrument means the protection of fundamental rights will always prevail over any other interest (e.g. innovation). But if different fundamental rights face each other, their application will require checks and balances to ensure that some are not safeguarded at the expense of others. For example, rules restricting deepfakes creation must be balanced with the freedom of expression. Any limitation under consideration will need to be set against the principles of legality, necessity and proportionality. According to the European Parliament, the legislative response needed in order to maintain a healthy democratic debate should both protect information and not compromise

²⁷⁵ M. VEALE AND F. Z. BORGESIU, *op. cit.* (voy. note 237), pp.4-5.

²⁷⁶ P. GLAUNER, *op. cit.* (voy. note 232), p.7.

fundamental rights.²⁷⁷ These checks and balances justify why the unacceptable risks category found under Article 5 AI Act should be well-defined and not too extensive. Otherwise, the ban would be counterproductive as it could encroach upon other EU-protected fundamental rights.

6.4.4. Foreign intervention in European and national democratic processes

Article 2(4) makes it clear that the regulation should not apply to public authorities in third country nor international organisations where those authorities or organisations use the AI systems in the framework of international agreements for law enforcement and judicial cooperation. The EDPB and EDPS deplored that the proposal did not rely on international law enforcement cooperation. According to their joint opinion, the proposal fails to address the issue of high-risk AI systems originating in foreign countries upon which EU public authorities might rely for law enforcement purposes. But other than that, the scope of the instrument allows for a broad application. AI systems providers or users located in a third country, where the output produced is used in the Union, will be subjected to the regulation. Foreign intervention is also covered by other EU instruments. For example, the European Parliament issued in 2019 a Resolution on foreign electoral interference and disinformation in national and European democratic processes.²⁷⁸

6.4.5. Transparency

The problem regarding the opacity of AI systems is only briefly addressed in the draft AI Act. The latter requires to improve the transparency of political advertising and communication. The obligation of Article 52 and the obligation of Article 13(1) for high-risk AI systems only are both transparency obligations. But they introduce very different threshold of transparency.

Article 13(1) – *High Risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. (...)*

Article 52(1) – *Providers shall ensure that AI systems intended to interact with natural persons are designed and developed in such a way that natural persons are informed that they that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use.*

²⁷⁷ EUROPEAN PARLIAMENT, “Foreign influence operations in the EU”, Briefing, <[https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI\(2018\)625123_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI(2018)625123_EN.pdf)> last accessed 19 July 2021, p.7.

²⁷⁸Resolution (EP) 2019/2810 of the European Union of 10 October 2019 on foreign electoral interference and disinformation in national and European democratic processes, <https://www.europarl.europa.eu/doceo/document/TA-9-2019-0031_EN.html> last accessed 29 July 2021.

While Article 13 demands interpretability, Article 52 requires disclosure.²⁷⁹ The notion of transparency in the draft requires clarification. Moreover, simply communicating to individuals the AI-based nature of the systems does not solve the disinformation issue or the opacity issue for that matter. It does not debunk the misinformation, nor does it explain or educate on the artificial process underlying the interaction. The second point of the European Democracy Action Plan of the European Commission, which aimed at strengthening media freedom and media pluralism, pinpointed two areas for improvement when it comes to transparency: communication and the transparency of political advertisements. Against the prospect of elections manipulation, Article 52 would certainly improve the transparency of AI systems in respect of communication, but it would not render political advertising more transparent. The target audience would know that they are dealing with an AI-generated content but they would still not know why and based on which data there are being targeted.

7. Conclusion

Internet was originally created to facilitate a hyper-connected world where ideas, inventions, knowledge and other proceeds from human activity could circulate freely and help further common progress. But it has recently turned into a source polarisation and misinformation instead. The rise of disinformation online has raised concerns regarding the protection of a genuine democratic process. Misinformation has grown widely especially before and around democratic elections. Political strategies feeding off the spread of disinformation aim at speaking to voters' emotions, and at confusing and dividing them in order to influence the elections' result. Recent technology developments have seen the intensification of the use of AI in disinformation schemes. This "disinformation 2.0" is at its dawn of possibilities but it has already raised the alarm among political experts and technology experts. Deepfakes, targeted political campaigning and political bots all resort to AI to be more realistic or to reach a wider and more specific audience (or both). In the current EU political climate, disinformation mechanisms are already associated with the rise of extreme parties and the weakening of democracy. And the weaponisation of AI mechanisms holds the potential for a more extensive impact. Even though their effects are greater in developing countries and non-democratic or unstable states, EU members states are not immune from the threat.

²⁷⁹ A. KISELEVA, "Making AI's Transparency Transparent: notes on the EU Proposal for the AI Act", European Law Blog, 2021, <<https://europeanlawblog.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/>> last accessed 31 July 2021.

Whether it is through personalisation, interest-driven usage or large-scale diffusion, AI election manipulation tools threaten the free and fair election process. To protect a healthy democratic debate, the Union's legislative response must find a balance between innovation and the protection of fundamental rights. AI brings a lot to the table in terms of innovation, but the EU should not be blindsided. Fundamental rights should not be the big losers in the ongoing race for technological innovation with the United States and China.

The EU is well-aware of the danger. The current legislative ambition of the Union is to create legislative instruments that enable both a flourishing innovation drive and the safeguard of democratic rights. Efforts to contain AI disinformation online are currently scattered across several instruments. Within this framework, the 2021 draft AI Act is a first attempt towards the regulation of AI systems. To be effective, the legislation will require a clear articulation with other instruments such as the GDPR and the future DSA. But the AI Act proposal does introduce new elements against AI elections manipulation. The draft sets out the very first harmonised definition of AI. Its risk-based approach also creates a scale of risks ranging from prohibited AI practices to lower risks practices. This scale leaves a leeway for innovation and allows for the protection of other coexisting fundamental rights (including the right to freedom of expression and the right to freedom of press). If introduced in its current form, the AI Act would address and regulate AI election manipulation tools under its Article 52, which imposes a minimal transparency obligation. This means that AI systems must be developed or designed in such a way that natural persons are at the very least aware that they are interacting with an AI system.

Although the issue of AI election manipulation tools finds its way into the scope of the draft AI Act, it appears from the analysis of the proposal that it has not been entirely written with these types of systems in mind. Despite the realistic danger and the all-important democratic implications, Article 5 AI Act, introducing a ban on certain AI systems practices, does not discuss at any point the possibility of redressing a collective/societal harm. This mere notion would have already broadened the possibilities for tackling elections manipulation. The proposal is of course still going through the ordinary legislative process (art. 294 TFEU) and can still be modified by the Council and the European Parliament.

There are many EU (and national) instruments discussing AI and misinformation techniques online. Undeniably, there is a need to develop a healthier democratic dialogue given the growing challenge of disinformation. But this must go through the creation of a comprehensive legal framework that is

also safeguarding legal certainty. We see candidates and governments using hate speech and fear on social media to convert these feelings into ballot results. Voters are bombarded with information online whether true, false, verified or not. Unaware of the extent of the profiling they are subjected to and lacking the necessary transparency to comprehend the inner working of algorithms, (targeted) voters become easier to influence. The constant broadcast of very simple and specific ideas may influence their feelings and ultimately impact their behaviour in upcoming elections.

“Knowledge is power” but it should be “knowledge” in the sense of understanding and questioning the information as well as having access to plural information sources. Yet, election manipulation is not an issue which calls for a “*one-size-fits-all*” solution. In a world where deception becomes an objective and a norm, communication and information are indeed a first step, but they cannot be the only one. As D. Citron advocates, there is a necessity to have a multi-methods solution including regulation, market action and education. It needs a systemic reflection on how education can meet with clear regulation and market responses (such as fact-checking devices). It requires that the Union and its member states invest in resources to engage with and understand both the mechanisms of disinformation and the technology supporting it. General and vague regulations or prohibitions are not viable solutions in the long run, even if the aim is drafting future-proof legislation. Time will tell where the AI Act fits in this spectrum.

II. Bibliography

Cour. Eur. D.H., *Ahmet Yildirim c. Turkey*, 18 December 2012, req. n° 3111/10.

ECHR, *Bowman v. The United Kingdom*, 19 February 1998.

ECHR, *Haralambie v. Roumanie*, 27 October 2009.

ECHR, *Satakunnan Markkinapörssi Oy and Satamedia Oy v. Finland*, 27 June 2017.

Charter of Fundamental Rights of the European Union of 18 December 2000, 2000/C 364/01.

Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union 2012/C 326/1.

Convention no. 108 for the Protection of Individuals with regard to Automatic Processing of Personal Data of 1 October 1981.

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) amending certain union legislative acts of 21 April 2021, COM (2021) 206 final.

Proposal for a Regulation of the European Parliament and the Council on a Single Market for Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Resolution (EU) 2019/2810 of the European Parliament of 10 October 2019 on foreign electoral interference and disinformation in national and European democratic processes, <https://www.europarl.europa.eu/doceo/document/TA-9-2019-0031_EN.html> last accessed 29 July 2021.

Resolution 2002/46 from the Commission on Human Rights relating to further measures to promote and consolidate democracy, Office for the High Commissioner for Human Rights, <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjRoI-0m5XwAhXABmMBHbvQDw0QFjABegQIAxAD&url=http%3A%2F%2Fap.ohchr.org%2Fdocuments%2FE%2FCHR%2Fresolutions%2FE-CN_4-RES-2002-46.doc&usq=AOvVaw2ZKWqdaYAM699P8OwkVGeL> last accessed 18 April 2021.

ABOKHODAIR N. et al., *Dissecting a social botnet: growth, content and influence in twitter*, Procedure of 18th ACM CSCW, 2015, <<https://arxiv.org/abs/1604.03627>> last accessed 3 July 2021.

AJDER H., PATRINI G., CAVALLI F. & CULLEN L., *The State of Deepfakes: Landscape, Threats and Impact*, Rapport Deepttrace, 2019.

AMBROZIAK A., “Sondaż OKO.press. Homofobia nie działa. Większość Polaków odrzuca opowieść o groźnej ideologii”, OKO.press, 2020, <<https://oko.press/sondaz-wiekszosc-odrzuca-opowiesc-o-groznej-ideologii-lgbt/>> last accessed 13 April 2021.

ANDERSON B. & HORVATH B., “The rise of the weaponized AI propaganda machine”, Scout, 2017, <<https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b>> last accessed 9 July 2021.

AYYUB R., *I was the Victim of a Deepfake Porn Plat Intended to Silence Me*, 2018, <https://www.huffingtonpost.co.uk/entry/deepfake-porn_uk_5bf2c126e4b0f32bd58ba316> last accessed 8 May 2021).

BAINES P. & SNOW N., *The SAGE handbook of propaganda*, SAGE publications Ltd., 2020.

BAYER J. et al., *Disinformation and propaganda – impact on the functioning of the rule of law in the EU and its Member States*, Study Requested by the LIBE committee, <[https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU\(2019\)608864](https://www.europarl.europa.eu/thinktank/en/document.html?reference=IPOL_STU(2019)608864)> last accessed 20 July 2021.

BOICHAK O. et. al., “Not the Bots You Are Looking For: Patterns and Effects of Orchestrated Interventions in the US and German Elections”, *International Journal of Communication*, 15, 2021, pp. 814-839, <<https://ijoc.org/index.php/ijoc/article/view/14866>> last accessed 17 August 2021.

BRACHTEN F. et al., *Strategies and Influence of Social Bots in a 2017 German state election: a case study on Twitter*, Australasian Conference on Information Systems, 2017, Hobart, <https://www.researchgate.net/publication/320564398_Strategies_and_Influence_of_Social_Bots_in_a_2017_German_state_election_-_A_case_study_on_Twitter> last accessed 20 July 2021.

BRADSHAW S. & HOWARD P. N., “The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation”, Computational Propaganda Research Project, Oxford Internet Institute, 2019, <<https://demtech.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf>> last accessed 20 July 2021.

BRELAND A., “The Bizarre and Terrifying Case of the “Deepfake” Video that helped bring an African Nation to the Brink”, Mother Jones, 15 March 2019, <<https://www.motherjones.com/politics/2019/03/deepfake-gabon-ali-bongo/>> last accessed 16 June 2021.

BROWN N.I., “Deepfakes and the Weaponization of Disinformation”, *Virginia Journal of Law & Technology*, 23(1), 2020, <https://static1.squarespace.com/static/5e793709295d7b60295b2d29/t/5f0b7e4e0f4f095f7e70613f/1594588750542/v23i1_1-Brown.pdf> last accessed 31 July 2021.

CADWALLADR C. & GRAHAM-HARRISON E., “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach”, The Guardian, 2018, <<https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>> last accessed 5 July 2021.

CADWALLADR C., “Facebook’s role in Brexit and the threat to democracy”, Session 1, TED2019 Bigger Than Us, 15 April 2019, <https://www.ted.com/talks/carole_cadwalladr_facebook_s_role_in_brexit_and_the_threat_to_democracy> last accessed 15 June 2021.

CADWALLADR C., “Google, democracy and the truth about internet search”, The Guardian, 2016, <<https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook>> last accessed 30 June 2021.

CADWALLADR C., “The great British Brexit robbery: how our democracy was hijacked”, The Observer, 7 May 2017, full article on <<http://davelevy.info/wp-content/uploads/2015/03/great-british-brexite-robbery.pdf>> last accessed 1 July 2021.

CAMBRIDGE DICTIONNARY, <<https://dictionary.cambridge.org/fr/dictionnaire/anglais/propaganda>> last accessed 18 June 2021.

CARDON D., *La démocratie Internet : promesses et limites*, Paris, Seuil, 2010.

CBINSIGHTS, “AI 100: The Artificial Intelligence Startups Redefining Industries”, 2021, <<https://www.cbinsights.com/research/report/artificial-intelligence-top-startups/>> last accessed 10 July 2021.

CHANDER S. & JAKUBOWSKA E., *EU’s AI law needs major changes to prevent discrimination and mass surveillance*, EDRI, 2021, <<https://edri.org/our-work/eus-ai-law-needs-major-changes-to-prevent-discrimination-and-mass-surveillance/>> last accessed 1 August 2021.

CHANNEL 4 NEWS INVESTIGATIONS TEAM, “Exposed: Undercover secrets of Trump’s data firm”, Channel 4 News, 2018, <<https://www.channel4.com/news/exposed-undercover-secrets-of-donald-trump-data-firm-cambridge-analytica>> last accessed 9 July 2021.

CHEN A. & POTENZA A., “Cambridge Analytica’s Facebook Data Abuse shouldn’t get credit for Trump”, The Verge, 2018, <<https://www.theverge.com/2018/3/20/17138854/cambridge-analytica-facebook-data-trump-campaign-psychographic-microtargeting>> last accessed 3 July 2021.

CHESNEY B. & CITRON D., “Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security”, *California Law Review*, 107, 2019, <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954#> last accessed 3 July 2021.

CHESNEY R. & CITRON D., “Disinformation on Steroids: The Threat of Deep Fakes”, Council on Foreign Relations, 2018, <<https://www.cfr.org/report/deep-fake-disinformation-steroids>> last accessed 3 March 2021.

CHRISTIANO T., « Democracy », in *The Stanford Encyclopedia of Philosophy* (ed. E.N. ZALTA), Fall 2018 Edition, <<https://plato.stanford.edu/archives/fall2018/entries/democracy/>>.

CHUNG L., “‘Lifelong impacts’: image-based abuse skyrockets online amid pandemic”, 2020, <<https://www.smh.com.au/national/nsw/lifelong-impacts-image-based-abuse-skyrockets-online-amid-pandemic-20201201-p56jpu.html>> last accessed 3 May 2021.

COHENDET M. A., « Une crise de la représentation politique », *Presses Universitaires de France*, 2004/2 (18), pp. 41-61, <<https://www.cairn.info/revue-cites-2004-2-page-41.htm>> last accessed 30 July 2021.

COMMISSION EUROPENNE, *Tackling Online Disinformation: an European Approach*, Communication de la Commission au Parlement européen, au Conseil, au Comité économique et social européen et au Comité des régions, 2018.

Commission guidance on the application of Union data protection law in the electoral context, a contribution from the European Commission to the Leaders' meeting in Salzburg of 19-20 September 2018 COM/2018/638 final.

CONFESSORE N., "Cambridge Analytica and Facebook: The Scandal and the Fallout So Far", The New York Times 2018, < <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>> last accessed 5 July 2021.

DAHL R. A., "Democratic polities in Advanced Countries: Success and Challenge", in *New Worldwide hegemony. Alternatives for change and social movements*, 2004, pp.51-70, < <http://biblioteca.clacso.edu.ar/ar/libros/hegeing/Dahl.pdf>> last accessed 30 May 2021.

DAMIANI J., "A Voice deepfake was used to Scam a CEO out of \$243,000", <<https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/?sh=d85064622416>> last accessed 22 February 2021.

DE COOREBYTER V., *Des pratiques démocratiques de crise*, Les @analyses du CRISP en ligne, 2015, <https://www.crisp.be/crisp/wp-content/uploads/analyses/2015-09-17_ACL-de_Coorebyter_V-2015-Des_pratiques_democratiques_de_crise.pdf> last accessed 3 May 2021.

DIGITAL, CULTURE, MEDIA AND SPORT COMMITTEE OF THE HOUSE OF COMMONS, "Disinformation and 'fake news': Final Report", Eighth Report of Session 2017-19, 2019, <<https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf>> last accessed 7 July 2021.

DIGITAL, CULTURE, MEDIA AND SPORT COMMITTEE OF THE HOUSE OF COMMONS, Oral evidence: Fake News, questions to the witness Brittany Kaiser, former Director of Program Development at Cambridge Analytica, 2018, video of the meeting: <<https://www.parliamentlive.tv/Event/Index/e5ae6255-c88e-4e62-bbf4-9c0c18ba7b6b>> last accessed 11 July 2021. Written evidence of the meeting: <<http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/digital-culture-media-and-sport-committee/fake-news/oral/81592.pdf>> last accessed 11 July 2021.

EASLEY D. & KLEINBERG J., *Networks, Crows, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, 2010, <<https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book.pdf>> last accessed 18 July 2021.

ECHR, Guide to the Case-Law of the European Court of Human Rights, Data Protection, 2020, <<https://rm.coe.int/guide-data-protection-eng-1-2789-7576-0899-v-1/1680a20af0>> last accessed 20 July 2021.

EDPB-EDPS, Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 2021, <https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf> last accessed 19 July 2021.

EKDALE B. & TULLY M., “African Elections as a Testing Ground: Comparing Coverage of Cambridge Analytica in Nigerian and Kenyan Newspapers”, *African Journalism studies*, 40(4), 2019, pp.27-43, DOI: 10.1080/23743670.2019.1679208.

ELKIT J. & SVENSSON P., “The Rise of Election Monitoring: What makes elections free and fair?”, *Journal of democracy*, 8(3), 1997.

EUROPEAN COMMISSION, “European democracy Action Plan: Making democracies stronger”, Factsheets, 2020, <https://ec.europa.eu/info/sites/default/files/edap_factsheet8.pdf> last accessed 3 July 2021.

EUROPEAN COMMISSION, *Commission reports on 2019 European elections: fostering European debates and securing free and fair elections*, Press release, 2019, <https://ec.europa.eu/commission/presscorner/detail/en/IP_20_1123> last accessed 1 June 2021.

EUROPEAN COMMISSION, Tackling Online Disinformation: A European Approach, COM/2018/236 Final, 2018, <<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0236&from=EN>> last accessed 10 July 2021.

EUROPEAN COMMISSION, The performance framework for the EU budget, under the 2021-2027 multiannual financial framework, volume II, Annex, 2021, <<https://op.europa.eu/en/publication-detail/-/publication/1d4ad9c8-d23b-11eb-ac72-01aa75ed71a1/language-en/format-PDF/source-search>> last accessed 14 July 2021.

EUROPEAN COMMISSION, White Paper on Artificial Intelligence: a European approach to excellence and trust, consultation, 2020, <<https://digital-strategy.ec.europa.eu/en/consultations/white-paper-artificial-intelligence-european-approach-excellence-and-trust>> last accessed 14 July 2021.

EUROPEAN DATA PROTECTION SUPERVISOR, Opinion 3/2018 on online manipulation and personal data of 19 March 2018, <https://edps.europa.eu/sites/default/files/publication/18-03-19_online_manipulation_en.pdf> last accessed 15 July 2021.

EUROPEAN PARLIAMENT, “Artificial intelligence, data protection and elections”, 2019, <[https://www.europarl.europa.eu/RegData/etudes/ATAG/2019/637952/EPRS_ATA\(2019\)637952_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2019/637952/EPRS_ATA(2019)637952_EN.pdf)> last accessed 14 June 2021.

EUROPEAN PARLIAMENT, “Foreign influence operations in the EU”, Briefing, <[https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI\(2018\)625123_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2018/625123/EPRS_BRI(2018)625123_EN.pdf)> last accessed 19 July 2021.

EUROPEAN PARLIAMENT, “Legislative Train 06.2021, A Europe fit for the digital age: proposal for a regulation on a European Approach for Artificial Intelligence”, <<https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence>> last accessed 11 July 2021.

EUROPEAN PARLIAMENT, “Parliamentary question E-001309/2021: answer given by Mr. Breton on behalf of the European Commission,”, 2021, <https://www.europarl.europa.eu/doceo/document/E-9-2021-001309-ASW_EN.html> last accessed 25 July 2021.

EUROPEAN PARLIAMENT, “Parliamentary questions: answer E-001309/2021 given by Mr. Breton on behalf of the European Commission”, 2021, <https://www.europarl.europa.eu/doceo/document/E-9-2021-001309-ASW_EN.html> last accessed 10 July 2021.

ERCAN S. A. & GAGNON J-P., “The Crisis of Democracy: Which Crisis? Which Democracy?”, *Democratic Theory*, 1(2), 2014.

EXTINCTION REBELLION, « La vérité sur le Covid-19 et la Crise écologie : un discours par Sophie Wilmès », <<https://www.extinctionrebellion.be/fr/dire-la-verite>> last accessed 20 February 2021.

FAGAN K., “A viral video that appeared to show Obama calling Trump a ‘dips—’ shows a disturbing new trend called ‘deepfakes,’” *Insider*, 2018, <<https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4?r=US&IR=T>> last accessed 22 February 2021.

FERRARA E. et al., “The Rise of Social Bots”, *Commun ACM*, 59(7), 2016, <<http://dx.doi.org/10.1145/2818717>> last accessed 24 July 2021.

GLAUNER P., “An Assessment of the AI Regulation Proposed by the European Commission”, preprint to appear in 2022 in *The Future Circle of Healthcare: AI, 3D, Printing, Longevity, Ethics and Uncertainty Mitigation* (eds. S. EHSANI et al.), 2022 <<https://arxiv.org/pdf/2105.15133.pdf>> last accessed 10 July 2021.

GOLDZWEIG R. & BRADY M., *Deepfakes: How prepared are we? Multi-stakeholder perspectives and a recommendations roadmap*, Democracy Reporting International, 2020, <<https://democracy-reporting.org/wp-content/uploads/2020/11/2020-11-Deepfakes-Publication-No-2-Web-file-1.pdf>> last accessed 10 March 2021.

GONZALEZ R.J., “Hacking the citizenry? Personality profiling, ‘big data’ and the election of Donald Trump”, *Anthropology Today*, 33(3), 2017, <https://doi.org/10.1111/1467-8322.12348>.

GOODFELLOW I. J. et al., *Generative Adversarial Nets*, Neural Information Processing Systems Conference Paper, 2014.

GOVERN T., *The rise of right-wing populism in Europe and the United States: a comparative perspective*, Friedrich Ebert Stiftung, 2016, <http://dc.fes.de/fileadmin/user_upload/publications/RightwingPopulism.pdf> last accessed 13 April 2021.

HABGOOD-COOTE J., “Stop talking about fake news!”, *Inquiry*, 62 (9-10), 2018, pp.1033-1065, <<https://www.tandfonline.com/doi/abs/10.1080/0020174X.2018.1508363?journalCode=sinq20>> last accessed 2 February 2021.

HAO K., *Deepfake porn is ruining women’s lives. Now the law may finally ban it*, MIT Technology Review, 2021, <<https://www.technologyreview.com/2021/02/12/1018222/deepfake-revenge-porn-coming-ban/>> last accessed 29 June 2021.

HENLEY J., “Court overturns referendums as voters were poorly informed... in Switzerland”, *The Guardian*, 2019, <<https://www.theguardian.com/world/2019/apr/11/switzerland-court-overturns-referendum-as-voters-were-poorly-informed>> last accessed 12 June 2021.

HERMET G., « Les choix fondateurs », in *La Démocratie* (ed. G. HERMET), Paris, Flammarion, 1997.

HOLLAND S. et. al., “The Dataset Nutrition Label: a Framework to Drive Higher Data Quality Standards”, in *Data Protection and Privacy: Data Protection and Democracy* (eds. D. HALLINAN et al.), Hart Publishing, 2020, <<https://media.bloomsburyprofessional.com/rep/files/9781509932740sample.pdf>> last accessed 19 July 2021.

HOUSE INTELLIGENCE COMMITTEE, *National Security Challenges of Artificial Intelligence, Manipulated Media, and Deepfakes*, 116th Congress, 2019, <<https://www.congress.gov/event/116th-congress/house-event/109620>> last accessed 3 June 2021.

HOWARD P. N., WOOLLEY S. & CALO R., “Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration”, *Journal of Information Technology & Politics*, 15(2), 2018, pp.81-93, DOI: 10.1080/19331681.2018.1448735.

IMPERVA THREAT RESEARCH LAB, “8th Annual Bad Bots Report”, <<https://www.imperva.com/resources/resource-library/reports/bad-bot-report/>> last accessed 26 July 2021.

JOHANNÈS F., « Comment le Rassemblement national est devenu le premier parti de la génération des 25-34 ans », *Le Monde*, <https://www.lemonde.fr/politique/article/2021/04/05/le-rassemblement-national-premier-parti-des-25-34-ans_6075574_823448.html> last accessed 31 July 2021.

KENSINGER E.A., “Negative emotions enhances memory: Behavioural and Neuroimaging Evidence”, *Sage Journal*, 16(4), pp.213-218, 2007.

KHAN R. & DAS A., *Build Better Chatbots: A complete Guide Getting Started with Chatbots*, 1st ed., Apress, 2018, <<https://doi-org.proxy.bib.ucl.ac.be:2443/10.1007/978-1-4842-3111-1>> last accessed 25 July 2021.

KISELEVA A., “Making AI’s Transparency Transparent: notes on the EU Proposal for the AI Act”, *European Law Blog*, 2021, <<https://europeanlawblog.eu/2021/07/29/making-ais-transparency-transparent-notes-on-the-eu-proposal-for-the-ai-act/>> last accessed 31 July 2021.

LANDEMORE H., *La démocratie représentative est-elle réellement démocratique*, La vie des idées, Interview of N.URBINATI & B. MANIN by H. LANDEMORE, <https://booksandideas.net/IMG/pdf/20080307_landemore.pdf> last accessed 2 June 2021.

LANGE S., “Plato and Democracy”, *The Classical Journal*, 34(8), 1939, <https://www.jstor.org/stable/3291135?seq=1#metadata_info_tab_contents>.

LATONERO M., *Governing Artificial Intelligence: Upholding Human Rights & Dignity*, *Data&Science*, 2018, <<https://apo.org.au/sites/default/files/resource-files/2018-10/apo-nid196716.pdf>> last accessed 14 July 2021.

LE MONDE AVEC REUTERS, « En Pologne, des milliers de manifestants après l’arrestation d’une militante LBGT », *Le Monde*, 2020, <<https://www.lemonde.fr/international/article/2020/08/09/en->

pologne-des-milliers-des-manifestants-apres-l-arrestation-d-une-militante-lgbt_6048524_3210.html> last accessed 29 April 2021.

LECUYER Y., *Le droit à des élections libres*, Conseil de l'Europe, « Hors Collection », Strasbourg, France, 2014, <<https://www-cairn-info.proxy.bib.ucl.ac.be:2443/le-droit-a-des-elections-libres--9789287177599.htm>>.

LESAFFER R., *European Legal History: a cultural and political perspective*, Cambridge, Cambridge University Press, 2009.

LOCKWOOD JR. T.C., “Polity, Political Justice and Political Mixing”, *History of Political Thought*, 27(2), 2006, pp.207-222, <https://www.jstor.org/stable/26222194?seq=6#metadata_info_tab_contents> last accessed 5 June 2021.

MAC S., “There are dozen ways to order a coffee. Why do dumb bots only allow one?”, Medium, 2017, <<https://medium.com/assist/theres-a-dozen-ways-to-order-a-coffee-why-do-dumb-bots-only-allow-one-27230542636d>> last accessed 24 July 2021.

MADDOCKS S., “‘A Deepfake Porn Plot Intended to Silence Me’: exploring continuities between pornographic and ‘political’ deep fakes”, *Porn Studies*, 7(4), 2020, DOI: 10.1080/23268743.2020.1757499.

MANHEIM K. & KAPLAN L., “Artificial Intelligence to Privacy and Democracy”, *The Yale Journal of Law and Technology*, 21, <https://yjolt.org/sites/default/files/21_yale_j.l._tech._106_0.pdf> last accessed 2 August 2021.

MANHEIM K. & KAPLAN L., “Artificial Intelligence to Privacy and Democracy”, *The Yale Journal of Law and Technology*, 21, <https://yjolt.org/sites/default/files/21_yale_j.l._tech._106_0.pdf> last accessed 2 August 2021.

MANRIQUE M., « Réseaux sociaux et medias d'information », *L'Harmattan « Confluence Méditerranée »*, 79, 2011, <<https://www.cairn.info/revue-confluences-mediterranee-2011-4-page-81.htm>> last accessed 13 accessed 2021.

MASSA P., “Trust It Forward: Tyranny of the Majority or Echo Chambers?” In *The Reputation Society: how online opinions are reshaping the offline world* (eds. H. MASUM & M. TOVEY), Cambridge, MIT Press, 2012.

MATIN N., “Online predators spread fake porn of me: here’s how I fought back”, TEDTalk, 2017, <https://www.ted.com/talks/noelle_martin_online_predators_spread_fake_porn_of_me_here_s_how_i_fought_back> last accessed 5 May 2021.

MATSUSAKA J. G., *Let the People Rule: How Direct Democracy Can Meet the Populist Challenge*, Princeton University Press, 2020.

MILLER F., “Aristotle's Political Theory”, in *The Stanford Encyclopedia of Philosophy* (ed. E. N. ZALTA), Winter 2017 Edition, <<https://plato.stanford.edu/archives/win2017/entries/aristotle-politics/>> last accessed 17 November 2020.

MOSSE C., *Regard sur la démocratie*, Paris, Perrin, 2013, pp.7-12.

MUELLER B., *The Artificial Intelligence Act is a Threat to Europe's Digital Economy and Will Hamstring the EU's Technology Sector in the Global Marketplace*, Centre for Data Innovation, Press Release, 2021, < <https://datainnovation.org/2021/04/the-artificial-intelligence-act-is-a-threat-to-europes-digital-economy-and-will-hamstring-the-eus-technology-sector-in-the-global-marketplace/>> last accessed 12 July 2021.

NATGUNANATHAN I. et al., *Protection of Privacy in Biometric Data*, IEEE Access, 2016, <<https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=7420576>> last accessed 20 July 2021.

NATIONAL DEMOCRATIC INSTITUTE, “Disinformation and Electoral Integrity: A Guidance Document for NDI Elections Programs”, 2019, <https://www.ndi.org/sites/default/files/Disinformation%20and%20Electoral%20Integrity_NDI_External_Updated%20May%202019%20%281%29.pdf> last accessed 10 June 2021.

NDULO M. & LULO S., “Free and fair elections, violence and conflict”, *Cornell Law Faculty Publications*, 51, 2010, <<https://scholarship.law.cornell.edu/cgi/viewcontent.cgi?article=1185&context=facpub>> last accessed 3 March 2021.

NENADIC I., “Unpacking the « European approach » to tackling challenges of disinformation and political manipulation”, *Internet Policy Review*, 8(4), 2019, <<https://policyreview.info/pdf/policyreview-2019-4-1436.pdf>> last accessed 15 August 2021.

NGUYEN T.T. et. al., *Deep Learning for Deepfakes Creation and Detection*, Cornell University, 2019, <<https://arxiv.org/pdf/1909.11573.pdf>> last accessed 13 February 2021.

NIX A., “The Power of Big Data and Psychographics”, Conference during the 2016 Concordia Annual Summit, see the video here: <<https://www.youtube.com/watch?v=n8Dd5aVXLCc>> last accessed 17 July 2021.

NOUJAIM J. & AMER K., “The Great Hack”, Documentary, 2019.

NYABOLA N., *Digital Democracy, Analogue Politics: How the Internet Era is Transforming Politics in Kenya*, London, ZED Books, 2018.

OHCHR, Chapter 23: Monitoring Human Rights in the Context of Elections, in: OHCHR, *Manual on human rights monitoring*, 2011, Professional Training Series n°7, <<https://www.ohchr.org/Documents/Publications/Chapter23-MHRM.pdf>> last accessed 10 June 2021.

PARIS B. & DONOVAN J., *Deepfakes and Cheap Fakes: the Manipulation of Audio Visual Evidence, Data & Society*, 2020, <https://datasociety.net/wp-content/uploads/2019/09/DS_Deepfakes_Cheap_FakesFinal-1-1.pdf> last accessed 10 March 2021.

PASQUALE F., “Chapter 5: Machine Judging Humans”, in *New Laws of Robotics: defending human expertise in the age of AI*, Cambridge, Massachusetts & London, The Belknap Press of Harvard University Press, 2020.

POSETTI J. et al., *Online violence Against Women Journalists: A global Snapshot of Incidence and Impacts*, UNESCO, 2020, <<https://www.icfj.org/sites/default/files/2020-12/UNESCO%20Online%20Violence%20Against%20Women%20Journalists%20-%20A%20Global%20Snapshot%20Dec9pm.pdf>> last accessed 15 June 2021.

POSSETTI J. & MATTHEWS A., *A short guide to the history of 'fake news' and disinformation: a learning module of journalists and journalism educators*, International Center for Journalists, 2018, <https://www.icfj.org/sites/default/files/2018-07/A%20Short%20Guide%20to%20History%20of%20Fake%20News%20and%20Disinformation_ICFJ%20Final.pdf> last accessed 28 February 2021.

Procedure 2021/0106/COD COM (2021) 206 for the Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) amending certain union legislative acts, <https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX:52021PC0206#2021-04-22_ADP_byCOM> last accessed 14 July 2021.

RASHBROOKE M., “3 ways to upgrade democracy in the 21st century”, TEDTALK, <https://www.ted.com/talks/max_rashbrooke_3_ways_to_upgrade_democracy_for_the_21st_century> last accessed 20 April 2021.

RATHI R., *Effect of Cambridge Analytica's Facebook ads on the 2016 US Presidential Elections*, towards data science, 2019, <<https://towardsdatascience.com/effect-of-cambridge-analyticas-facebook-ads-on-the-2016-us-presidential-election-dacb5462155d>> last accessed 24 April 2021.

RONFAUT L., « “Deepfakes”: avant la présidentielle, le grand faux dans l'inconnu », Libération, 2021, <https://www.liberation.fr/economie/economie-numerique/deepfakes-avant-la-presidentielle-le-grand-faux-dans-linconnu-20210514_ZDSIMQUA5BH3TNJBC2AJ5I4O3E/> last accessed 5 July 2021.

ROYLE S., “Deepfake porn images still give me nightmares”, BBC, 2021, <<https://www.bbc.com/news/technology-55546372>> last accessed 4 July 2021.

SABONI A. & BLANGERO A., *Opening the algorithm's black box and understand its outputs*, 4th International Conference on Electrical and Information Technologies, IEEE, 2020, <<https://ieeexplore.ieee.org/abstract/document/9113174>> last accessed 25 July 2021.

SCHWEMER S. F. et al., *Legal AI Systems in the EU's proposed Artificial Intelligence Act*, Legal AIIA, 2021, <<http://ceur-ws.org/Vol-2888/paper8.pdf>> last accessed 24 July 2021.

SEF: DEVELOPPMENT AND PEANCE FUNDATION, “We need to bring the human back into the digital conversation”,

Interview with Nanjala Nyabola on social media and its role in Kenya sef: insight 2, 2019, <<https://www.sef-bonn.org/en/events/dresden-forum-for-international-politics/2019.html>> last accessed 5 July 2021.

SHAO C. et al., “The spread of low-credibility content by social bots”, *Nature Communication*, 9, 2018, <<https://doi.org/10.1038/s41467-018-06930-7>> last accessed 28 July 2021.

SHU K., WANG S., LEE D., LIU H. et. al., “Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements”, in *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities* (eds. K. SHU, S. WANG & H. LIU), Spring International Publishing, 2020.

SIMONITE T., “What Happened to the Deepfake Threat to the Election?”, *Wired*, 2020, <<https://www.wired.com/story/what-happened-deepfake-threat-election/>> last accessed 3 June 2021.

SINTOMER Y., « Le tirage au sort à travers l’histoire : une domestication du hasard ? », in *Petite histoire de l’expérimentation démocratique. Tirage au sort et politique d’Athènes à nos jours* (ed. Y. SINTOMER), Paris, La Découverte, Poche/Essais, 2011, <<https://www-cairn-info.proxy.bib.ucl.ac.be:2443/petite-histoire-de-l-experimentation-democratique--9782707170149-page-39.htm>>.

SOLON O. & LAUGHLAND O., “Cambridge Analytica closing Facebook data harvesting scandal”, *The Guardian*, 2018, <<https://www.theguardian.com/uk-news/2018/may/02/cambridge-analytica-closing-down-after-facebook-row-reports-say>> last accessed 4 July 2021.

SPECIAL COUNSEL ROBERT S. MUELLER III, “Report on the Investigation into Russian Interference in the 2016 Elections”, 1(1-5), 2019, <<https://www.justice.gov/archives/sco/file/1373816/download>> last accessed 30 July 2021.

SPECIAL COUNSEL ROBERT S. MUELLER III, *Report on the Investigation into Russian Interference in the 2016 Elections*, 1(1-5), 2019, <<https://www.justice.gov/archives/sco/file/1373816/download>> last accessed 30 July 2021.

STAAL J., *Propaganda Art in the 21st Century*, MIT Press, Cambridge, 2019.

SUSSER D., ROESSLER B. & NISSENBAUM H., “Technology, autonomy, and manipulation”, Alexander von Humboldt Institute for Internet and Society, *Internet Policy Review*, 8(2), Berlin, 2009, pp. 1-22, <<http://dx.doi.org/10.14763/2019.2.1410>> last accessed 13 July 2021.

TIDEY A., “Poland to give money to “LGBT free” town denied EU funding”, *Euronews*, 2020, <<https://www.euronews.com/2020/08/19/poland-to-give-money-to-lgbt-free-towns-denied-eu-funding>> last accessed 13 April 2021.

TIMAN T. & MANN Z., “Data Protection in the Era of Artificial Intelligence: Trends, Existing Solutions and Recommendations for Privacy-Preserving Technologies”, in *The Elements of Big Data Value* (eds. E. CURRY et al.), Springer International Publishing, 2021.

TINLAND F., « Hobbes, Spinoza, Rousseau et la Formation de l’idée de démocratie comme mesure de la légitimité du pouvoir politique », *Revue Philosophique de la France et de l’Etranger*, 175(2), 1988, pp.195-222, <https://www.jstor.org/stable/pdf/41093850.pdf?casa_token=SpMAL5-iTF0AAAAA:t43x9tRPOfmZ7rSULKRLCV38z33Sd0EczAkyPXGu_vKwHbqJjMwZR5HfPCFaA-IN59aYg3FQBjYgiTfSXxOPW5WXnmNjOgi22LKAxHUSq765sogo4T6> last accessed 27 May 2021.

TOURAINÉ A., “What is democracy?”, the UNESCO Courier, <<https://en.unesco.org/courier/novembre-1992/what-democracy>> last accessed 1 April 2021.

UNESCO, *Journalist, Fake news and Disinformation: Handbook for Journalism, Education and Training*, UNESCO Series on Journalism Education, 2018.

UNITED NATIONS HUMAN RIGHTS OFFICE, *UN experts call on India to protect journalist Rana Ayyub from online hate campaign*, 2018, <<https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23126&LangID=E>> last accessed 25 April 2021.

UNITED NATIONS, *Freedom of Expression and Elections in the Digital Age, Mandate of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, 2019, Research Paper 1/2019.

UNITED NATIONS, Guidance notes from the Secretary-General on democracy, <https://www.un.org/democracyfund/sites/www.un.org.democracyfund/files/un_sg_guidance_note_on_democracy.pdf> last accessed 3 February 2021.

UNITED NATIONS, *Report of the Special Rapporteur on the promotion and the protection of the right to freedom of opinion and expression, Frank La Rue*, General Assembly, Human Rights Council, 2014.

URBINATI N., “Representative democracy and its critics”, in *The Future of Representative Democracy* (eds. S. ALONSO, J. KEANE, W. MERKEL), Cambridge University Press, 2011.

VEALE M. & BORGESIU F. Z., “Demystifying the Draft EU Artificial Intelligence Act”, *Computer Law Review International*, 22(4), 2021, <<https://arxiv.org/abs/2107.03721>> last accessed 25 July 2021.

VON DER BURCHARD H., “Belgian socialist party circulated ‘deepfake’ Donald Trump video”, Politico, 2018, <<https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/>> last accessed 14 June 2021.

VOSOUGHI S., ROY D. & ARAL S., The spread of true and false news online, *MIT initiative on the digital economy*, 2018, pp. 1146-1151, <<https://ide.mit.edu/sites/default/files/publications/2017%20IDE%20Research%20Brief%20False%20News.pdf>> last accessed 2 April 2021.

WAGNER T.L. & BLEWER A., “The Word Real is No Longer Real: Deepfakes, Gender and the Challenges of AI-Altered Video”, *Open Information Science*, 3(1), 2019.

WALKER S., “Polish President issues campaign pledge to fight LGBT ideology”, The Guardian, 12 June 2020, <<https://www.theguardian.com/world/2020/jun/12/polish-president-issues-campaign-pledge-to-fight-lgbt-ideology>> last accessed 10 April 2021.

WARDLE C. & DERAKSHAN H., *Information Disorder: Towards an interdisciplinary framework for research and policy making*, Council of Europe Report, 2017.

WOLTERSTORFF N., *Understanding liberal democracy: Essays in political philosophy*, Oxford University Press, 2012.

WOOLLEY S., “Automating power: social bot interference in global politics”, *First Monday*, 21(4), 2016, <<https://doi.org/10.5210/fm.v21i4.6161>> last accessed 19 July 2021.

WOOLLEY S., *The Reality Game: How the next wave of technology will break the truth and what can we do about it*, London, Endeavour, 2020.

ZALC J., BECUWE N. & BURUIN A., *The 2019 post-electoral survey: have European elections entered a new dimension*, Eurobarometer Survey 91.5 of the European Parliament, A Public Opinion Monitoring Study, 2019.

III. Glossary of terms

Algorithms: finite sequence of formal rules (logical operations and instructions) making it possible to obtain a result from the initial input of information. This sequence may be part of an automated execution process and draw on models designed through machine learning.²⁸⁰

Artificial intelligence (AI): set of scientific methods, theories and techniques whose aim is to reproduce, by a machine, the cognitive abilities of human beings. (...) The tools identified in this [article] are developed using machine learning methods, i.e. “weak” AIs.²⁸¹

Big data: large sets of data from mixed sources (e.g. open data, proprietary data and commercially purchased data). For data derived from judicial activity, big data could be the combination of statistical data, records of business software connections (application logs), court decisions’ databases, etc.²⁸²

Black box: system or a software whose internal functioning is not understandable. In machine learning, a black box model is created directly from data by an algorithm, meaning that humans, even those who design them, cannot understand how variables are being combined to make predictions.²⁸³

Bots: online automated software program, bots can be made to communicate with both computer systems and human users.²⁸⁴

Deepfake: amalgamation of the words “deep learning” and “fake”; these videos or audios are resorting to artificial intelligence, specifically deep learning software to manipulate videos or audios of a subject using other extant imagery or video of that subject.²⁸⁵

Deep learning: branch of machine learning which is itself a branch of AI. Deep learning networks make use of artificial neural networks which are based on communication mechanism in human

²⁸⁰ European ethical Charter on the use of Artificial Intelligence in judicial system and their environment of the European Commission for the Efficiency of justice (CEPEJ) adopted on the 3-4 December 2018, p.69.

²⁸¹ European ethical Charter, *op. cit.* (voy. note 280), p.69.

²⁸² European ethical Charter, *op. cit.* (voy. note 280), p.70.

²⁸³ C. RUDIN & J. RADIN, “Why are we using black box models in AI when we don’t need to? A lesson from an explainable AI competition”, *Harvard Data Science Review*, 1(2), 2019, p.2.

²⁸⁴ S. WOOLLEY, *op. cit.* (voy. note 82), p.203.

²⁸⁵ S. WOOLLEY, *op. cit.* (voy. note 82), p. 204.

brain to learn from sets of information or data which are not structured, contrary to machine learning.²⁸⁶

Generative adversarial network: type of machine-learning modelling tool like initial, or training data. This systematic technology is particularly useful in generating realistic-seeming images or videos of a person that are, in fact, manipulated using other extant images of that person.²⁸⁷

Information warfare: operation aiming at manipulating the information available to opponents in order to gain a military advantage by influencing their opinions, attitudes and actions.²⁸⁸

Machine learning: branch of artificial intelligence tools that work to autonomously catalogue structured data sets by following patterns and other indicators.²⁸⁹

Profiling: as defined by the GDPR, it is a form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preference, interests, reliability, behaviour, location or movements.²⁹⁰

Psychological warfare: military use of propaganda to influence the mind an opponent. These techniques are also called psyops (short for "psychological operations").²⁹¹

Turing test: method established by Alan Turing in 1950 which assesses whether an artificial intelligence is capable of thinking like a human being. During this test, the same question will be asked to both a computer and a human being responder. The interrogator will then have to assess which one is, despite the intelligence behaviour, non-human.²⁹²

²⁸⁶ S. WOOLLEY, *op. cit.* (voy. note 82), p.205.

²⁸⁷ S. WOOLLEY, *op. cit.* (voy. note 82), p.205.

²⁸⁸ J. ARO, "The cyberspace war: propaganda and trolling as warfare tools", *European View*, 15, 2016, pp.121-132, <<https://link.springer.com/content/pdf/10.1007/s12290-016-0395-5.pdf>> last accessed 20 July 2021, p.122.

²⁸⁹ S. WOOLLEY, *op. cit.* (voy. note 82), p.205.

²⁹⁰ Regulation (EU) 2016 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), art. 4(4).

²⁹¹ J. ARO, *op. cit.* (voy. note 288).

²⁹² A. M. TURING, "Computing Machinery and Intelligence", *Mind*, LIX (236), 1950, pp.433-460, <<https://academic.oup.com/mind/article/LIX/236/433/986238>> last accessed 3 August 2021.

Page deliberately left blank.

