

Smart society – “Fake analytica“ style?*

Bissera ZANKOVA

Media 21 Foundation, Sofia, Bulgaria

E-mail address: bzankova@gmail.com

Abstract

The article will focus on the problems generated by the information disorder including the recent “Analytica” scandal and how they imperil the creation of “smart society”. Essential for the existence of such society is smart citizenship which is under threat in the age of disinformation. Important questions are related to the quality of participation – one about the diffusion of fake news and hate speech in the web and another to counteracting them through various initiatives. As a framework of the study I shall use the European human rights standards and regulations which are indispensable to the establishment of a smart environment for all persons and communities. Against this backdrop the risks for participation and engagement through social media will be analyzed. The study will try to find answers to the following questions: do and to what extent social media trigger information disorder, how Facebook will change after “Analytica” (if at all) and how the users will change their participation habits and behaviour (if at all)? In the current research I shall use and enrich my previous investigation projects of civic journalism, blogs and protests in Bulgaria (in co-authorship with A. Dimova, 2013), of smart journalism in co-authorship with Zankova, Skolkay, Franklin, 2016, of smart citizenship, social networks and local referendums as a tool for expanding community democracy [1] and the most recent one on the effectiveness and efficiency of fact-checking and debunking organizations (in co-authorship with T. Pavleska, A. Skolkay, N. Ribeiro, A. Bechmann (forthcoming)). This interdisciplinary piece will be useful for both academics and practitioners and specifically for media and NGO specialists to get knowledge about the increased accountability of the media today, and about the conditions that will guarantee secure people’s involvement in an active and meaningful online participation and debate.

Keywords: information disorder, fake news, hate speech, social media, participation.

1. Introduction

Today our reliance on the Internet is ubiquitous and social networks promoting active communication have become indispensable to our daily activities. The Internet establishes a space for free expression of thought and opinion offering broader opportunities for diverse and alternative voices to be heard. “The increasingly prominent role of social networking services and other social media services also offer great possibilities for enhancing the potential for the participation of individuals in political, social and cultural life”, the Council of Europe **Recommendation CM/Rec(2012)4[2] of the Committee of Ministers states**. International organizations have already acknowledged the public service value of the Internet and the role of social networks in fostering pluralism and diversity and in facilitating social cohesion, participation and democracy in particular. (**Recommendation CM/Rec(2012)4 [2]**) Earlier in 2010 through its Recommendation **CM/Rec(2014)7 of the Committee of Ministers the Council of Europe passed the Charter on Education for Democratic Citizenship and Human Rights Education. Through this special instrument it provides guidance in the education**

*This article is based on the research carried out within COMPACT, Horizon 2020, EC project: Compact: from research to policy through raising awareness of the state of the art on social media and convergence. Project Number 762128, call: H2020-ICT-2016-2017, topic: ICT-19-2017.

for democratic citizenship “By equipping learners with knowledge, skills and understanding and developing their attitudes and behaviour, to empower them to exercise and defend their democratic rights and responsibilities in society...with a view to the promotion and protection of democracy and the rule of law” the formulated objective reads.**(Charter on Education for Democratic Citizenship)**

Nowadays democratic citizenship is a prerequisite to smart citizenship. It is related to the process of active involvement in governance at all levels and demonstrates keen attitude towards issues of public concern on a daily basis. From a political perspective, our desire is to build “smart democracy”, which is based on committed and enlightened citizenship. [1] "Smart citizens" refer to a group of people that take an open and informed stand in transforming the overall environment including cities into a better one. Citizens' collaboration in any undertaking is one of the essential characteristics of modern societies if they wish to develop and prosper. “The role of citizens as active stakeholders is becoming more and more important every day. A closer look at our daily routines reveals the huge importance that technology and social networks have taken on in our lives and we need to analyze how this has facilitated the creation of increasingly horizontal bridges” is the message of the Smart City Expo World Congress, held in Barcelona in November 2018.

To live in harmony, to accomplish one's goals and to contribute to the individual and social well being is a normal human desire. A smart environment can support such cravings to become real practice. “Smart society is one of these complex but promising concepts that remain poorly understood. This futuristic language sounds like a science fiction, but **smarter societies are already slowly emerging all around us**” **Krasnodebsci optimistically concludes (Krasnodebsci, 2018)** What has to be born in mind is that “technological change is never neutral” and it always involves ethical, environmental and political considerations”. “Any progressive step may have pitfalls and raise challenges to the human mind.” (Benjamin Le Roux in **Krasnodebsci 2018**)

Further to this the fast technological development to a great extent outpaces the changes in the social and legal perception of the power of technologies and their impact on society. The same much praised “smart conditions” giving impetus to innovative forms of social action and cooperation can generate frictions or sometimes serious problems with respect to human rights which may jeopardize the intelligent coexistence of communities and persons. In contrast to the bright perspectives some recent events representing examples of “information disorder” have put on trial our trust in the positive opportunities opened by the technological revolution. Three phenomena at least have posed serious questions before modern society and they need to be effectively addressed for the sake of its smart future.

In the next section I shall dwell on the most topical issues on the public agenda that have provoked a broad discussion about the controversial role of social media in our lives. The first one is about disinformation and its expansion on the net not without the contribution of the social networks.

2. Information disorder

In the digital world we may encounter a variety of communications problems each of different scope and impact. The “information disorder” creates serious dilemmas before users and groups. The rapid spread of fake news, propaganda and hybrid wars which particularly accelerate in times of political transformations or social upsurge preceding elections is one of the biggest challenges of our time. In 2016 “The Guardian” observing how fake information is distorting politics in many countries, named fake news “an insidious trend that’s fast becoming a global problem”[3]. This admonition is even more relevant today. At the beginning of November 2018 Twitter deleted more than 10,000 automated accounts that posted messages to stop people to go voting in the US elections and render themselves Democrats. The measures were taken after the party noticed the misleading publications on the social network, Reuters reported.

The number seems modest given the fact that Twitter has already erased millions of accounts they identified as responsible for sending disinformation before the US presidential elections in 2016 [4].

In Europe the EU citizens are troubled by the prospects of interference in upcoming European Parliament elections. A Eurobarometer survey revealing the expectations of the Europeans for the MEPs elections in May 2019 came to this alarming conclusion as a total of 61% of the interviewed were worried that elections might be affected by cyber attacks; 59% feared that people outside the Union and criminal groups might influence the elections and 67 % were concerned that their personal data online could be used to send targeted political messages to them [5].

These forebodings if come true will comprise the next malicious blow on the foundations of the democratic society as democracy presupposes government and civic responsibility are exercised by all citizens with full understanding of their role in the democratic processes, directly or through freely elected representatives.

2.1 *Mis-, dis - and mal information*

“Fake news” which is at the core of disinformation is a term that can mean different things, depending on the context. These may include false political content intermingled with propaganda but also rumours, gossip or generally the purposeful dissemination of information that is not checked, is not objective and in the worst cases is completely misleading. More often than in any other time we resort to the question whether all these expressions (that have already been known to the media) represent different forms of fake news and where the fluid line of truth and falsity lies. From a journalistic perspective verifying news has always been a challenging task and that is why professional journalism should abide by clear-cut rules demanding truth, objectivity and fairness of coverage summarized in the form of ethical codes, statements or declarations. Today journalism is undergoing deep transformations due to the intensive use of social media and the wide publishing of user generated content (citizen journalism). While the profession democratizes and opens to new players, **ethics** remain vital for the entrenching of good practices against human rights violations and abuse of public trust.

Fake messages and hoaxes can generate diverse repercussions - social, political, economic, digital. Bounegru et al. [6] argue that “fake news cannot be fully understood apart from its circulation online.” Therefore in order to identify how big the real problem is we have to take into account the reach of false information and the different modifications caused in the overall communication environment.

In 2018 the EC elaborated a definition of fake news which can be considered as a living tool following the dynamics of social development and communication. For the Commission fake news and disinformation are equal: “Disinformation - or fake news - consists of verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm.” [7]. However, practice (as usual) proves to be much more nuanced than definitions. In their seminal book “Information Disorder. Toward an interdisciplinary framework for research and policymaking” Wardle and Derakhshan [8] create a holistic framework for detecting and categorizing various forms of false information according to the purpose and level of dangerousness. We can speak of mis-information when false information is shared, but no harm is meant, of dis-information when false information is knowingly shared to cause harm and of mal-information when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere. Such comprehensive approach to defining various forms of false information may serve also as a basis to devising a range of measures that can be applied in each particular case.

Though the mildest form of false information, “misinformation can be very difficult to correct and may have lasting effects even after it is discredited. One reason for this persistence is the manner in which people make causal inferences based on available information about a given event or outcome. As a result, false information may continue to influence beliefs and attitudes even after being debunked if it is not replaced by an alternate causal explanation” [9]. Therefore the effects of false content can be long lasting and generally deleterious for the formation of enlightened public opinion, for taking sound decisions and for democratic values. Commenting on the launch of the report produced by the UK Commons’ digital, culture, media and sport committee on disinformation and “fake news” the chair, Damian Collins, stressed “we are facing nothing less than a crisis in our democracy based on the systematic manipulation of data to support the relentless targeting of citizens, without their consent, by campaigns of disinformation and messages of hate.” [10]. It is vivid that the negative effects of information disorder strike directly at the principles of our society. The phenomenon is having stronger impact now than ever before as it is easier for anyone to post and share any news or opinion online. The users receiving information do not always have the capacity to evaluate it and its sources with precision [11]. They are not the only to be blamed for the information disorder problems. The general impression is that Facebook (FB) consistently refuses to acknowledge its practical, moral or legal responsibilities. Even if we assume FB as the most popular social media does care about news it understands the term differently than many in a newsroom or the public at large. In FB’s eyes personal news also matters and in the combination of news of various sorts, rumours, subjective and perverted information can easily flourish. Subsequently the

overall quality of information can affect the quality of the debate and the goals pursued. Most dangerous is the political use of the methods of disinformation for the attacks on the mental condition of members of society.

Stephan Russ-Mohl, Director of the European Journalism Observatory (2018), argues that “Eastern European countries are a particularly good playing field for disseminating disinformation using new information technologies. Especially the Baltic States and Ukraine continue to be the targets of permanent Russian disinformation attacks” [12]. The dissemination of false stories can be a global process but in Eastern Europe its implications are the most forceful and intertwined with geopolitical projects. Bulgarians are among the Europeans who least trust the news on social media, according to data from a Eurobarometer survey. Two percent of the Bulgarians and five per cent of the rest of the EU only fully believe the news online [13]. According to the latest study on freedom of the media and trust in international institutions and international organizations in Bulgaria a large proportion of the respondents considered print media (37%), social networks (35%) and even television (20%) the largest distributors of false allegations (possibly multiple answers) [14]. In the fast changing environment beliefs and perceptions change quickly to one or another direction and we cannot be sure what data will be available in the future. Moreover surveys premised on different methodologies may also contradict each other. The information uncertainty can be the reason that some risks which are the results of information disorder may not be captured or counteractive efforts weakened.

The situation of information disorder affects human beliefs, ideas and social plans to the extent that they are devised in a perverted manner. Eventually disinformation can undermine citizens’ trust in the democratic principles and values and particularly in freedom of expression and the media. Thus our hopes that we pave the way to the building of a smart society underpinned by free and active citizenship can be totally lost.

In the next subsection I shall discuss another information disorder problem which is intimately related to human dignity and human rights and also puts at stake democratic participation - hate speech.

2.2 Hate speech

The low-cost, high-speed and immediate dissemination via social media platforms created an ideal scene for spreading hate speech. A growing number of groups has emerged devoted to human rights’ hostile ideas as homophobic, Islamophobic, anti-immigrant, anti-Semitic hate, misogyny, white supremacy, etc. The main problem related to online hate speech is whether its effects can be traced to real life events and unlawful actions. The findings so far are not encouraging. Hateful speech, even if it does not reach the threshold to turn into “incitement to violence”, can be detrimental to human relationships and reinforce negative biases, prejudice and stereotypes in the society. Gelashvili [15] referring to Waldron claims there are “two dangerous types of messages in hate speech that expose different groups to vulnerability.” The first (narrower message - B.Z.) is directed at the victims with the intention to stigmatize or ridicule them and to make them

feel excluded from the society. The second (broader message – B.Z.) can reach an overall social effect that is even more humiliating – “to insult victims, stereotyping them, for example, as terrorists, advocating the exclusion of them from society, denying them human rights, holding them accountable for the actions of the other members of the group, applying double standards etc” [15]. A third effect has a bearing on the fundamental principles of society - such conduct strikes at the heart of democratic citizenship based on equality and freedom and eventually puts at risk democracy.

As early as 2012 the Council of Europe in its already quoted Recommendation CM/Rec(2012)4 [2] of the Committee of Ministers to member States on the protection of human rights with regard to social networking services has pointed out that “the right to freedom of expression and information, as well as the right to private life and human dignity may also be threatened on social networking services, which can also shelter discriminatory practices” The instrument enlists specifically the risks which may arise from “lack of legal, and procedural, safeguards surrounding processes that can lead to the exclusion of users; inadequate protection of children and young people against harmful content or behaviours; lack of respect for others’ rights; lack of privacy-friendly default settings; lack of transparency about the purposes for which personal data are collected and processed.” (Recommendation CM/Rec(2012)4 [2]) Addressing hate speech on social media and countering its consequences is a difficult task as private entities – social media owners staying closer to the content circulated and holding considerable power over the flow of information and ideas online - pursue their own ends. The way in which social media giants have handled the issues pertaining to hate speech are not always acceptable especially from the standards of freedom of expression as every one of them applies its own less or more liberal policy of content publishing which is also not fully transparent and familiar to the public. Google, Facebook and Twitter have different policies regarding whether and what kinds of hate speech are permitted on their sites. Besides there have been serious accusations against Twitter for gender discriminatory practices from an Amnesty International Report. A survey published by the organization in 2018 called Twitter “a toxic environment for women” [16]. The document states that “harmful and negative gender stereotypes of women offline, as well as widespread discrimination against women rooted in patriarchal structures, manifest as violent and abusive tweets against some women on Twitter” [16]. Policies (if such) are also inconsistently implemented and can be hard for users to understand them. The public at large is not aware what guarantees of their rights are in place when communicating on social networks – either because of the non-transparent policies applied or due to their own ignorance about these matters.

As the EC reported since May 2016, Facebook, Twitter, YouTube and Microsoft have committed efforts to combating the spread of illegal hate speech online content in Europe through the adopted Code of Conduct [17]. The third monitoring round shows that the companies are fulfilling their commitment to remove the majority of illegal hate speech within 24 hours. However, the Commission admits there is still more to be done ahead. There is lack of regular communication with users on the issue and particularly “lack of systematic feedback to users.” [18]. There is also no sufficient clarity what standards are used to categorize certain forms of expression as hateful and discriminatory in order to be

removed. Many of the decisions made by the content removal teams at the social media organizations are hasty and do not take into account the rules protective of freedom of expression and the public interest in free speech. Current situation gives social media companies an enormous power to control what videos, text, images, etc. users may or may not post or access on social media sites.

It is a complex task to decide whether a specific message can be classified as illegal and, as such, whether it should or could legitimately be prohibited. Therefore in practice Facebook, Twitter and other social media control the everyday implementation of freedom of expression on social networks. From a more general perspective, any restriction on this freedom, whatever the objective it seeks to achieve, necessarily raises a series of legal questions related to the proper balancing of human rights. “International human rights law requires states to jointly protect and promote the rights to freedom of expression and the right to equality: one right cannot prevail over the other, and any tensions between them must be resolved within the boundaries of international human rights law” is the opinion of Article 19 [19]. International norms mandate particularly severe forms of ‘hate speech’: “incitement to genocide” and “advocacy to discriminatory hatred that constitute incitement to violence, hostility and discrimination” to be banned and in exceptional circumstances criminalized. However, the criteria under which certain forms of expression are prohibited have to be clear and proven in each specific case. On the other hand, states are also obliged to create an enabling environment for the exercise of the right to freedom of expression and the right to equality, and pursue the comprehensive protection of these rights within the implementation of their positive function. These are norms of universal significance that keep society together no matter what degree of smartness it has accomplished. They are valid at any level and with respect to all bodies and organizations. Any response in the form of any type of regulation of ‘hate speech’, intolerance and discrimination should be based on the human rights principles and should be considered ensuring the protection of these rights. It is widely known that the right to freedom of expression is not absolute. However, it can be limited in exceptional cases only and any limitation must conform to the strict requirements of foreseeability, legitimacy of aim and necessity in a democratic society.

Hate speech is a form of information disorder in social media that on an every day basis imperils basic human rights, weakens social cohesion and threatens the smart development of society. In addition to this hate speech compromises also free speech as a core right and value which is at the roots of democratic citizenship.

Another form of information disorder is the *Analytica data* scandal which has radically changed people’s attitude towards digital technology and its application. Apparently the latter can generate both unprecedented opportunities and catastrophic risks for society. In the following subsection I shall briefly sketch the most important human rights and rule of law issues stemming from this notorious case.

2.3 The “Analytica data” scandal

After FB disclosed that information from 50 million accounts on the social network was used without people’s permission by a company called “*Cambridge Analytica*”, a digital consultancy hired by the Trump presidential campaign, the world we live in is not the

same. Such piece of news has shocked even the most devoted supporters of the digital society and its benefits. This personality theft in cyber space supported the building of psychographics to influence political perceptions and voting behavior. Investigations by the Intercept, the Guardian/Observer, the New York Times, and undercover footage from the British Channel 4 network revealed the company crossed the line between “consumer persuasion to political manipulation” [20].

One of the violations in this case obviously pertains to the use of personal data without the permission of its owners. Another relates to the misuse of the same data for political purposes “to promote fear and distrust” [20]. It is paradoxical that privacy as a right helps avoid unwanted and potentially intrusive interference in personal affairs on the part of the state. In the case *Marckx./Belgium*, the European Court of Human Rights (ECtHR) stated ‘*the object of the Article [8] (of the European Convention on Human Rights – ECHR – B.Z.) is “essentially” that of protecting the individual against arbitrary interference by the public authorities*’. In the *Analytica* case the offender was not the state but a private company and it transpired that dangers for privacy were even greater nowadays as they could come from many more subjects than the state.

Data protection in some international acts is considered an extension of the right to privacy. However, it is a fundamental right on its own merits enshrined in Article 8 of the EU Charter of Fundamental Rights (CFA) being distinct from the respect for private and family life contained in Article 7 of the Charter. This feature distinguishes the EU Charter from other major human rights documents. For the European policy data protection is fundamental which has led to the adoption of the General Data Protection Regulation (GDPR) enforced on May 25 2018. The implementation of the regulation brings forth more complex issues regarding everybody’s personal accountability when collecting and processing personal data on the net. Though meant to ensure effective protection the act may appear complicated to comply with on a daily basis. Moreover, all active social networks’ users have to be fully aware of their role as data controllers. An interesting observation is that in a social network the ISP provides the profiles but the control over their functionality is exercised by the consumer [21]. The characteristic of a personal data administrator depends on the conscious choice of the consumer to offer content and to initiate operations with personal data. One can conclude that all individuals who professionally use social networks (politicians, journalists, socially active persons, sole traders and self-employed persons) must take account of their obligations under the regulation as personal data controllers. More specifically they take this role if they define the purposes and the means for the processing of personal data. [21] The “domestic exception” which holds when absolutely personal or domestic activities are performed still exists but it should be strictly interpreted (*ECJ Ryněš judgment (C-212/13)*). Individuals using their social networking profiles for both personal and professional activities will not be able to benefit from the exception and may fall within the scope of the regulation as personal data controllers. Such conclusion which is based on an in-depth legal analysis and the court jurisprudence factually evidences that with respect to data protection and privacy everyone who is thoroughly involved in social networks’ activities has higher obligations notwithstanding their profession [21].

Regarding its social repercussions the *Analytica* case could have had not only a disturbing effect on people's minds. It might have also made persons and communities look at social platforms with critical eyes and particularly at FB. As CNN Money points out: "All of this comes as Facebook is already getting questions about the long-term appeal of its platform, at least in the United States. The number of daily active users in the United States - a whopping 184 million - declined for the first time last quarter. The Cambridge Analytica scandal is likely to hasten user disenchantment with the network, sources inside FB acknowledged." [20] We can expect that users will more carefully scrutinize FB now as it will be viewed as a platform prone to manipulation by political groups, governments and lobbies. May be it is a moment of truth, when the sobering of the public enables it to reconsider what kind of environment social media create and what has to be done for its improvement.

The misuse of data from more than 50 million FB accounts have also drawn persons' attention to the more general but also more sensitive issue of the implementation of fundamental principles like rule of law and human rights in the digital world. The principle is that "the same rights that people have offline must also be protected online"[22]. However in the digital age, not everything concerning the implementation of norms depends on the legal frameworks. There is also a technological impact which can stay new and unexplored for a period of time. "The processes and algorithms that govern our online experiences are always biased to some extent but we never really know how, to what degree and whether at all they affect our human existence." Suzor claims [23] Access to the Internet is a precondition for the exercise of human rights and freedoms nowadays. The logical questions that follow are who and how governs the Internet and whether legal frameworks contain adequate guarantees for the principles indispensable to democracy in the digital age. Apparently their implementation related to human rights and rule of law is not the same as in the real world. In 1996 John Perry Barlow published a Declaration of the Independence of Cyberspace in which he declared enthusiastically that "we will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before" [24]. Regretfully to conclude this dream has not come true yet. The *Analytica* debacle created a good occasion to bring again to the fore fundamental themes as human rights, regulation and governance in the digital environment. A sub-issue is whether the Internet and social media in particular need more regulation and what type of regulation would be appropriate in order for to not stifle free debate. As Zuckerberg has reluctantly admitted: "I'm not sure we shouldn't be regulated," "There are things like ad transparency regulation that I would love to see." During the CNN interview Zuckerberg repeated also some of the technical steps FB is going to take the "*Cambridge Analytica*" data exploit wouldn't happen again. To investigate all apps that have access to large amounts of data, and restrict developers' data access even further, to make sure everyone knows whose data was affected by one of "these rogue apps". He also promised more transparency should apps stray from the company's terms in the future. Zuckerberg also added: "Facebook is going to build a tool that will allow users to determine whether their data was affected" [25]. He did not say a word about human rights, rule of law, transparency and public accountability and the apps policy could be the next mystification meant to exonerate the social network. Another important point is that all his ideas have to turn into concrete actions in no time and

become known to society as the public suspicion is that FB continues operating along the same opaque lines.

So far the pressure even the ban on the Internet usage put by the governments of some countries has been considered the greatest jeopardy for free and creative modern communication. Data misappropriation causing damages both to its owners and to the whole democratic system and especially to freedom of expression serves as a proof that the dangers flowing from the non-transparent and arrogant sometimes governance of private platforms is none the less menacing for a smart society. Such actions imperil not only normal access to information and ideas but also citizens' participation, social activism and mobilization online. The issue of regulation of social media is on the public agenda again as stakeholders are becoming more and more concerned with the falsehood, hate messages, and incitement to violence that populate cyberspace (and social media in particular). Citizens are increasingly demanding protection of their rights with respect to their larger engagement in diverse projects and initiatives on the net.

3. Discussion and conclusion

In this article I have started a discussion about the challenges in a world driven by fast-expanding technologies in which a smart society is expected to be built. They have led to the proliferation of social media which is a new social universe due to the interactive participation of persons, groups and communities. Nowadays social media facilitate communication, association and mobilization. Active involvement in social networking presupposes higher obligations for the parties. I have also recalled our cherished objective to accomplish smart democracy premised on active citizenship. The advantages created by the Internet were juxtaposed to at least three manifestations of information disorder we all have been witnessing recently. Phenomena like disinformation, hate and discriminatory speech online and personal data misappropriation challenge our trust in social networks as genuinely social and rights-fostering mechanisms. That is why I have raised the dilemma we currently face - are social media really shaping a smart society underpinned by enlightened and participatory citizenship or they create a monstrous one "Fake – Analytica" style? Such question can definitely provoke more queries than immediate clear answers.

The truth is nowadays social media are demonstrating not only their democratic potential but serious flaws in their operation. The negative implications of the latter may not amount always to cybercrime but in virtual reality categories are relative. In their article on contemporary cyberware Almeida, Doneda and Abreu [26] present examples of possible cybercrime that can develop into cyberwarfare. "Malicious pieces of code (viruses, trojans, root-kits, worms, bots, and spyware) and weaponized "zero days" can be deployed not only to perpetrate common cybercrime but also to engage in cyberwarfare. Identity theft, online scams and fraud, and theft of intellectual property or classified information usually fall under the first category — that of "common cybercrimes." Other cyber activities, depending on their scale, effects, originators, and targets, are sometimes characterized as a "cyber act of war." "There's no litmus test for the distinction between the two groups of malicious activities", the authors conclude [26]. This means that in cyberspace there can be fast transformations from one threat or violation to another which

can accelerate and this fact should not be ignored. Besides, all unlawful acts performed through the Internet even the mildest ones factually cause information disorder that threatens human rights, rule of law and democracy on a larger or on a smaller scale.

The first reaction that comes to mind regarding cyber crime and cyber offences is to improve the global Internet governance, to strengthen cooperation and human rights protection - a task engaging the efforts of a plethora of stakeholders on a national and global level. The next response which falls within the scope of our research here relates specifically to social media and how to ensure their proper functioning to the interest of society and democracy. If a subject is media it carries a public function, if a media is social it has to serve members of society abiding by certain rules. The three problems social media have engendered call for fast and adequate measures because they can lead to more damaging implications on social and community life. However, these reactions should not provide simple or piecemeal answers but smart solutions. Searching for suitable approaches some scholars explicitly point to the contextual and cultural dimensions of spreading hate speech and fake news online. Different interests can be involved in the process and speech contexts are not deprived of ambiguity. In order to understand properly digital discourse it is important to elucidate the practices of communication that go beyond the division line and through a historical, cultural, contextual and comparative analysis. The precision of the used terminology can also be discussed and reviewed as "hate speech", for instance does not have a universal international law definition and does not signify unconditionally what goes in or out of the scope of protected expression within any communication or tradition. The same inference can be related to mis-,dis- and mal-information and all the variations of fake news as in the digital environment boundaries are elusive.

Without ignoring the importance of the quality of information and exchange I think that the gist of the discussion about information disorder has a much wider dimension and it is about the fate of freedom of expression in the digital age and whether the Internet can function as a genuinely free space for debate unfettered by the censorship of corporations and governments. Some years ago the Internet was considered as the most liberal and creative environment, nowadays the Internet bears the burdens of a meticulously and pervasively regulated media space through norms, standards and codes.

As a starting point an in-depth study of human rights in the digital age seems more than necessary with a view of finding a suitable policy and legislative measures which on the one hand, can counteract risks and on the other, will not restrict human rights and free expression in particular. As Coccoli[27] argues "Indeed, the idea that the relationship between freedom of expression and ICT is capable of producing undoubted benefits to society is widely recognized.... The situation gets more complicated when it comes to the consideration that a traditional view of freedom of expression doesn't necessarily require the ability to share that expression. On the contrary, using ICT means that such self-expression is intended to be shared with a broader social community and causes a transformation from a basically private phenomenon into a mainly public one, making human rights extremely vulnerable." The author suggests that the first issue that needs to be examined is whether today's regulatory framework on human rights is sufficiently

suitable to guarantee freedom of expression in legal situations created by the new technologies[27] or has to be reinterpreted taking account of the principles of “indivisibility and interdependence of human rights in the light of digital innovations experienced by society” [27].

In addition Suzor[23] emphasizes that there is also a problem that “most of our theory and laws about constitutional restrictions on power apply only to nation states”. Transnational media companies exceed in power the latter but being private entities limitations on their conduct is hard to impose. However, we have to stress that businesses also have the obligation to protect human rights and this is espoused by the UN Guiding Principles on Business and Human Rights [28] and its European counterpart the **Recommendation CM/Rec(2016)3[29] of the Committee of Ministers to member States on human rights and business** “(though they are not legally binding and backed by an enforcement mechanism). “The responsibility to respect human rights is a global standard of expected conduct for all business enterprises wherever they operate” states the UN document [28]. This principle can be a salient basis for evaluating what is happening on the net and discussing in a multistakeholders’ format what should be done in the future.

In any media environment the exercise of the right to free expression can come into conflict with other human rights. However, in order to protect freedom of expression as a foundational right, principle and value in a democracy the effects of the technological expansion have to be studied carefully and approached in a manner that does not abuse free communication. Bottom up initiatives as the Wikitribune site that will fight fake news following the Wikipedia's model can be particularly useful and inspiring because they engage collective efforts and may ensure greater publicness. The encouragement of fact checking and debunking organizations in Europe is factually recognition of a new stakeholder among others on which the task to combat information disorder is set. **Fact-checkers** should establish an independent European network which will be a model of common working methods to exchange best practices supported by a secure European online platform on disinformation. Besides these organizations should improve their transparency which will contribute to the overall transparency of anti-disinformation activities [7]. However, the status of fact checking and debunking organizations has to be clarified and significantly strengthened. Alarming signals come from FB as journalists working as factcheckers have tried to cancel a controversial media partnership with the social network, saying the company has ignored their concerns and failed to use their expertise to combat misinformation. Moreover results achieved are minimal and mutual trust is gone as the platform is fueling the same type of propaganda (anti-Semitic against Soros – B.Z.) as fact checkers regularly debunk [30].

Having in mind the complex environment regulatory responses to the cases of information disorder are not always the best ones. They can be passed under the pressure of the day or of petty political interests, can suppress open and free debate on the Internet and can embolden authoritarian regimes to limit legitimate speech under the guise of the defense against extremism. Besides, such solutions do not always provide the necessary human rights and due process guarantees though they raise good arguments in this direction. In this respect one could recall the reactions to the German social media law

requiring social media companies to remove obviously illegal content, including hate speech and other postings within 24 hours after receiving a notification or complaint, and to block other offensive content within seven days, considered by bloggers and activists a “dangerous abridgement of free speech rights” [31]. The EU approaches are cautious and place the responsibility for the removal of illegal expression on the social platforms. On 1 March 2018, the Commission issued a Recommendation on measures to effectively tackle illegal content online.(Commission Recommendation (EU) 2018/334 [32]) Online platforms need to exercise a greater responsibility in content governance and be proactive pursuing a range of measures including **clearer 'notice and action' procedures, more efficient tools and proactive technologies, stronger safeguards to ensure fundamental rights, special attention to small enterprises and closer cooperation with authorities. Regarding human rights protection the European Commission calls for** effective and appropriate guarantees especially when automated tools are used “in full respect of fundamental rights, freedom of expression and data protection rules.”The cooperation resulting in the adoption of the already mentioned Code of Conduct resembles a co-regulatory initiative, however, as Article 19 rightly observes the Code was neither subjected to a public discussion, nor non-governmental organizations participated largely in its preparation which is a shortcoming from the perspective of civil society involvement and active citizenship [33].Similar comments can be made with respect to the new Code of Practice on Disinformation issued by the EC but relying on voluntary self-regulations of online platforms, leading social networks, advertisers and agencies which envisages actions on a host of things, from transparency and political advertising to closing fake accounts, to try and head off online election meddling ahead of the spring 2019 European elections. By and large regulatory attempts like these can be effective if they complement measures related to the improvement of platforms’ governance aiming more particularly at accomplishing greater accountability and transparency in their operation and decision-making but this again is a goal that the owners themselves should pursue.

Self-regulation which can be considered the most appropriate approach to regulate any media including social media is also not the panacea to any problem. Self-regulation can be dispersed, ineffective and depends on the will of the self-regulated. In countries where the state has traditionally played a powerful role in the media field, and where there is stricter content regulation through legislation, there can be reluctance to put in force a system of self-regulation. In these countries media ethics commissions that administer self-regulatory codes can be decorative structures only without real authority in society. In addition self-regulation of social media is different from self-regulation of traditional media comprising a set of technical and ethical rules. Speculating on self-regulation (platform algorithms) the Portuguese Civil Institute of the Self-Discipline of Commercial Communication (ICAP) stressed that platform self-regulation can flow from a variety of scenarios such as legal, technical or improvement of the business model. The process should be premised on accountability, transparency and public criticism as public involvement is crucial in this respect. The choice of an approach can rely on several factors: the nature of the algorithm, the context in which it exists or a risk analysis [11]. Such examples are thought provoking and illustrate that novel approaches vis-a-vis social media should be searched for due to their unique nature. They can be either completely

new or devised on the basis of reassessed traditional models with a view to creating really workable solutions in the digital environment.

Last but not least the problem of media literacy and digital empowerment of users not only as consumers but as citizens should be constantly on the public agenda. In a smart society knowledge and participation prove fundamental for the accomplishment of individual and social objectives. A central issue nowadays is whether we are aware of our human rights as individuals and citizens and how they are protected against abuses online. Pursuing social networking and exchange we have to exercise not only our rights but fulfill also our obligations to protect the rights of others and especially their private space and personal data because the structure and principles of operation of social media is such that every one who joins in is responsible for the creation and dissemination of content. Being alert to the behaviour of social media companies and ISPs does not preclude one's personal responsibility as user and citizen.

In the ideal case social media giants should not only abstain from violating human rights and due process on the web but through their policies take a pro-active stand for the enhancement of these rights. A fruitful idea can be instead of investing in the technical methods for removal of ill content, filtering and take down procedures and apps social platforms to invest in projects fostering human rights education, democratic citizenship and participation which as a positive obligation can produce better and long-term results.

A human rights based policy comprises the universal basis underpinning the appropriate conduct of all stakeholders in the digital environment and constitute the necessary foundation for constructing a smart society. An open discussion about the concrete steps and methods can facilitate this process and alleviate the consequences of the “Fake-(Hate)-Analytica” phenomena.

References

[1]Zankova, B.(2018)[Smart citizens for smart cities: the role of social media for expanding local democracy \(the case of local referendums in Bulgaria\)](#). In:[Smart Cities and Regional Development \(SCRD\) Journal](#), Universul Academic Publishing House, vol. 2(2), pages 19-33, June.

[2][CM/Rec\(2012\)4](#) of the Committee of Ministers to member States on the protection of human rights with regard to social networking services. (Adopted by the Committee of Ministers on 4 April 2012 at the 1139th meeting of the Ministers' Deputies. Retrieved 1 January 2019 from https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=09000016805caa9b.

[3]Connolly,K., Chrisafis,A., McPherson,P., [Kirchgaessner,S.](#), [Haas,B.](#), Phillips,D.,[Hunt, E.&Safi, M.](#) (2016). Fake news: an insidious trend that's fast becoming a global problem. With fake online news dominating discussions after the US election, Guardian correspondents explain how it is distorting politics around the world, 2 Dec 2016. Retrieved 1 January 2019 from <https://www.theguardian.com/media/2016/dec/02/fake-news-facebook-us-election-around-the-world>.

[4]OnAir. Twitter deleted more than 10,000 accounts trying to influence the US elections.The tweets have tried to make people vote. (inBulgarian). (Twitter изтри над 10 хил. акаунта, опитващи да повлияят на изборите в САЩ

[5]**FactorBG**. EUcitizensfearinterferenceintheEuropeanParliamentelections. 67% are concerned that their personal data online can be used to make targeted political messages, 26.11.2018. (inBulgarian).(Гражданитена ЕС сеопасяватотнамеса в изборитезаЕвропейскипарламент.67%

сазагрижени,

челичните им данни онлайн може да се използват за отпращане на целенасочени политически послания). Retrieved 1 January 2019 from <https://www.investor.bg/evropa/334/a/grajdanite-na-es-se-opasiavat-ot-namesa-v-izborite-za-evropejski-parlament-272555/>.

[6] Bounegru, L., Gray, J., Venturini, T. & Mauri, M. (eds.) (2017). A Field Guide to Fake News: A Collection of Recipes for Those Who Love to Cook with Digital Methods. Public Data Lab, Research Report.

[7] European Commission Press release 2018 Tackling online disinformation: Commission proposes an EU-wide Code of Practice. Retrieved 1 January 2019 from http://europa.eu/rapid/press-release_IP-18-3370_en.htm.

[8] Wardle, C. & Derakhshan, H. (2017). Information Disorder. Toward an interdisciplinary framework for research and policymaking. Strasbourg: Council of Europe.

[9] Nyhan, B. & Reifler, J. (2015). Displacing Misinformation about Events: An Experimental Test of Causal Corrections, *Journal of experimental political science*, volume 2, issue 1, April 2015, pp. 81 – 93. Retrieved 1 January 2019 from www.cambridge.org/core/journals/journal-of-experimental-political-science/article/displacing-misinformation-about-events-an-experimental-test-of-causal-corrections/69550AB61F4E3F7C2CD03532FC740D05.

[10] The Guardian view on the fight against fake news: neutrality is not an option. Editorial. 29 July 2018. Retrieved 1 January 2019 from <https://www.theguardian.com/commentisfree/2018/jul/29/the-guardian-view-on-the-fight-against-fake-news-neutrality-is-not-an-option>.

[11] DESINFORMAÇÃO EM ELEIÇÕES. Desequilíbrios acelerados pela tecnologia. Instituto de tecnologia e equidade (2018). Retrieved 1 January 2019 from https://gallery.mailchimp.com/9b54dd7f6a0bdac9e2cd34ddd/files/030f71ec-011b-4535-9696-4e3d8a172da4/ITEE_Pesquisa.pdf.

[12] Russ-Mohl, S. (2018). *Fake News” and Disinformation in Eastern Europe*, April 19 2018. Retrieved 1 January 2019 from <https://en.ejo.ch/specialist-journalism/fake-news-and-disinformation-in-eastern-europe>.

Smart City Expo World Congress 2018. Retrieved 1 January 2019 from <http://www.smartcityexpo.com/bg/web/smartcityexpo/blog/-/blogs/1642649>.

[13] Bulgarians are among the Europeans who least believe in the news. *Dariknews World (2018). (in Bulgarian)* (Българитесредевропейците, които най-малко вярват в новините). Retrieved 1 January 2019 from <https://dariknews.bg/novini/sviat/bylgarite-sa-sred-evropejcite-koito-naj-malko-viarvat-v-novinite-2084037>.

[14] Antonova, V. (2018). Citizens in Bulgaria do not trust the media. 63% of them are dependent on someone according to an Alfa Research and Konrad Adenauer Foundation research., 21.11.2018. (in Bulgarian).

[15] Gelashvili, T. (2018). Hate speech on social media: implications of private regulation and governance gaps. Retrieved 1 January 2019 from <http://lup.lub.lu.se/luur/download?func=downloadFile&recordId=8952399&fileId=8952403>

[16] Amnesty International. Toxic Tweeter – a toxic place for women. (2018). Retrieved 1 January 2019 from <https://www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter>

[17] European Commission. Code of Conduct on countering illegal hate speech online: First results on implementation. Factsheet. December 2016. Retrieved 1 January 2019 from <http://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupDetailDoc&id=29738&no=1>.

[18] European Commission Press release. Counteracting illegal hate speech online – Commission initiative shows continued improvement, further platforms join, Brussels, 19 January 2018. Retrieved 1 January 2019 from http://europa.eu/rapid/press-release_IP-18-261_en.htm.

[19] Article 19a. (2018). Responding to “hate speech”. Comparative overview of six European countries. Retrieved 1 January 2019 from <http://europeanjournalists.org/mediaagainstthate/wp-content/uploads/2018/02/Final-compilation-off-regional-research-digital.pdf>

[20] Doug, P. (2018). Cambridge Analytica Scandal: A Threat to Democracy and Facebook’s Future, 20 March 2018. Retrieved 1 January 2019 from https://sandiegofreepress.org/2018/03/cambridge-analytica-scandal-a-threat-to-democracy-and-facebooks-future/#.W_2EAjFoS9I.

[21] Sulev, G. (2014). Social media consumers as personal data administrators. (in Bulgarian) – In: Society and law, N5, 2014. (Потребителите на социални мрежи като администратори на лични данни, Общество и право, №5, 2014 г.).

- [22]UN. General Assembly. Human Rights Council. Twentieth session. Agenda item 3. Promotion and protection of all human rights, civil, political, economic, social and cultural rights, including the right to development, 29 June 2012.
- [23]Susor, N.(2016).Who governs the Internet? The rule of law in a decentralized world. Sept 16 2016.Retrieved 1 January 2019 from. <https://digitalsocialcontract.net/governing-the-internet-the-rule-of-law-in-decentralized-regulation-626bab07c73e>.
- [24]Barlow, J. P (1996). A Declaration of the Independence of Cyberspace. Retrieved 1 January 2019 from <https://www.eff.org/cyberspace-independence>.
- [25]Musil,S.(2018).Zuckerberg 'sorry' for data scandal, open to regulation. Facebook's CEO tells CNN he regrets his company's handling of the Cambridge Analytica scandal. 21 March 2018. Retrieved 1 January 2019 from <https://www.cnet.com/news/facebook-ceo-mark-zuckerberg-apologizes-for-cambridge-analytica-data-scandal/>.
- [26]Almeida, V.A.F., Doneda.D.&Abreu. J.de S.(2017). Cyberwarfare and digital governance. Retrieved 1 January 2019 from https://www.researchgate.net/publication/314164227_Cyberwarfare_and_Digital_Governance.
- [27]Coccoli,J.(2017).The challenges of new technologies in the implementation of human rights: an analysis of some critical issues in the digital era. In: Peace Human Rights Governance,1(2), 223-250. Retrieved 1 January 2019 from http://phrg.padovauniversitypress.it/system/files/papers/2017_2_4.pdf.
- [28]UN Guiding principles on business and human rights. Implementing the United Nations “Protect, Respect and Remedy” Framework (2011). - Retrieved 1 January 2019 from https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf.
- Waldron, J. (2012). The harm in hate speech. Cambridge, Mass., London: Harvard University Press, 2012, pp. 2-3.
- [29]Recommendation **CM/Rec(2016)3 of the Committee of Ministers to member States on human rights and business** (Adopted by the Committee of Ministers on 2 March 2016 at the 1249th meeting of the Ministers' Deputies). Retrieved 1 January 2019 from https://search.coe.int/cm/Pages/result_details.aspx?ObjectId=09000016805c1ad4.
- [30]Levin, S. (2018) The Guardian. 'They don't care': Facebook factchecking in disarray as journalists push to cut ties, 13 December 2018. Retrieved 1 January 2019 from <https://www.theguardian.com/technology/2018/dec/13/they-dont-care-facebook-fact-checking-in-disarray-as-journalists-push-to-cut-ties>.
- [31]Asgeirsson, E.(2017).German social media law threatens free speech, April 10 2017. Retrieved 1 January 2019 from <https://www.humanrightsfirst.org/blog/german-social-media-law-threatens-free-speech>.
- [32]Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online. Retrieved 1 January 2019 from <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32018H0334&from=GA>.
- [33]Article 19b.(2018).Self-regulation and 'hate speech' on social media platforms.Retrieved 1 January 2019 from https://www.article19.org/wp-content/uploads/2018/03/Self-regulation-and-%E2%80%98hate-speech%E2%80%99-on-social-media-platforms_March2018.pdf.