The Battle to End Fake News: A Qualitative Content Analysis of *Facebook* Announcements on how it combats disinformation

Petros Iosifidis (corresponding author) and **Nicholas Nicoli**

Published online in ***International Communication Gazette,*** October 2019

Actual Publication in Issue 82, Number 1, February 2020.

Accepted Version:

ABSTRACT

The recent spread of online disinformation has been profound and has played a central role in the growth of populist sentiments around the world. Facilitating its progression have been politically and economically motivated culprits who have ostensibly taken advantage of the digital freedoms available to them. At the heart of these freedoms lie social media organisations that only a few years earlier techno-optimists were identifying as catalysts of an enhanced digital democracy. In order to curtail the erosion of information policy reform will no doubt be essential (Freedman, 2018). The UK's *Disinformation and 'fake news' Report* (DCMSC of the House of Commons, 2019) and *Cairncross Review* (Cairncross, 2019), and the *European Commission's Report on Disinformation* (2018) are three recent examples seeking to investigate how precisely such reform policy might be implemented. Just as important is how social media organisations take on more responsibility and apply

self-regulating mechanisms that stifle disinformation across their platforms (something the aforementioned reports identify). Doing so will go a long way in restoring legitimacy in these significant institutions. *Facebook* (which includes *Instagram* and *Whatsapp*), is the largest social media organisation in the world and must primarily bear the burden of this responsibility. The purpose of this paper is to offer a descriptive account of *Facebook's* public announcements regarding how it tackles disinformation and fake news. Based on a qualitative content analysis covering the period November 16th 2016 – March 4th 2019, this paper will set out some groundwork on how to hold social media platforms more accountable for how they handle disinformation.

KEYWORDS

Facebook, social media, communications policy, disinformation, fake news, post-truth, misinformation, qualitative content analysis, populism

INTRODUCTION

Liberal democracies, political systems with free elections and basic liberties, are increasingly pushed toward illiberal, populist tendencies (Isaac and Rose, 2018). These are tendencies of a divisive and systemic global new order (Fukuyama, 2018) in which the role of social media has been significant (see Flew 2017, for a discussion on where media technologies might be leading us culturally, economically, politically and socially). Social media, once thought of as harbingers of digitized public spheres, are increasingly used by partisan tribal communities to selectively expose and share their own populist ideals.

Populist sentiments are often elicited through online disinformation by actors wishing to take advantage of the changing political and social landscape. These include extreme ideological groups, foreign agents, or actors simply wishing to make profits from unsuspected users. Online disinformation is currently one of the most pressing challenges of the digital age (Bennett and Livingston, 2018). Policy reform effectively overseeing online disinformation will no doubt alleviate the erosion of social media and curb escalating cynicism toward news media content (Freedman, 2018). Indeed, Mark Zuckerberg, Facebook's CEO and founder, has admitted that the Internet needs new rules curtailing the power internet companies have over speech (Zuckerberg, 2019). Across different countries and continents, policy reform is already being drafted or amended. Just as important is how social media organisations take on more responsibility and apply self-regulating mechanisms that stifle disinformation across their platforms.

The purpose of this paper is to offer a descriptive account of Facebook's public announcements on how it tackles online disinformation. Based on a qualitative content analysis we will set out some groundwork on Facebook's inner workings of

how it fights back against online disinformation. To consider this, we start by unpacking the conceptual and theoretical framework relating to populism and the crisis of democracy before turning our attention to policy issues, online disinformation and Facebook.

POPULISM AND THE CRISIS OF DEMOCRACY

Donald Trump's rise to political power in the US, the upsurge of the Alternative für Deutschland (AfD) in Germany, the increased influence of the French National Rally and the UK's Brexit vote have taken many by surprise. Indeed, as Choy suggests (2018: 752), "reliance on self-reported data, the potential bias of using landline numbers, the lack of appropriate sampling frames, and a spiral of silence effects", might explain current limitations in predicting both votes and ideological attitudes. Yet the rise of (mainly right-wing) populism has been a threat to Western democracies at least since the 1990s and has intensified after the 2008 global financial crisis (Norris and Inglehart, 2019). The different manifestations of right-wing populism share a common feature: they attack or even compromise the core elements of democratic societies such as the separation of powers, protection of minorities, or the rule of law (Fitzi, Mackert and Bryan, 2018). Left-wing populism has also been on the rise (see Mouffe, 2018). This has been particularly the case in the Mediterranean countries of Greece, Spain and Portugal that were hit hard by the banking crisis and were imposed neo-liberal austerity measures. Populisms of the political left include SYRIZA in Greece, which went into power in 2015, Podemos in Spain, Jean-Luc Melenchon, a far-left candidate who did very well in the French presidential election of 2017, Bernie Sanders in the US, but also the Jeremy Corbyn movement in the UK. The reasons for the electoral successes of populist parties vary

widely, but Inglehart and Norris (2016) attempted to group them into two camps: first, the economic inequality perspective, which is in fact the most widely held view of mass support for populism; second, the cultural backlash thesis, which suggests that support for populist parties is not a purely economic phenomenon but it is in large part a reaction against progressive cultural shifts. The authors propose that cultural values, combined with several social and demographic factors, provide the most consistent explanation for voting support for populist parties.

Flew and Iosifidis illustrate in the introduction of this *International Communication Gazette* special issue what populism is within liberal democracies. Moffitt (2016) argues that populism is exercised across various political and cultural contexts and calls for a rethinking of the concept, which may not be based just on the classic divide between 'the people' and 'the elite', to incorporate its shifting relationship to political representation, its global nature, and its reliance on new media technologies. This latter point may help us understand why populism has seemingly spread so rapidly across the globe at a time when media pervades political life and social media networks are increasingly used as tools by populist politicians to get into power. In an elegant attempt to clarify the notion of populist philosophy in the European context, Mudde (2007) suggested that the philosophy is a loose set of notions that have in common three core features: anti-establishment (populism reflects ordinary people as against the 'corrupt' establishment); authoritarianism (populism depicts resentment with existing political authorities, vested economic interests and big media firms); and nativism (populist discourse promotes nativism or xenophobic nationalism). The volume by Fitzi, Mackert and Bryan (2018) offers critical views on the debate on populism from the perspectives of political economy and the analysis of critical historical events, the links of analyses of populism with

social movement mobilisation, the significance of 'superfluous populations' in the rise of populism and an analysis of the exclusionary character of populism from the perspective of the theory of social closure.

As may be gathered from the above brief discussion populism has a broad meaning and has assumed a multitude of forms. This article is organised around the key themes that are pertinent to contemporary populism and its links to today's media landscape. The existing literature, while growing, does not as yet sufficiently consider the media-centred shifts occurring across politics and how the latter is increasingly reshaped due to new and social media's influence. We attempt to understand populist leaders' nuanced adaption of social media networks and strategies as a core factor in the spread of the populist phenomenon. By doing so, we ask a controversial question: does populism represent a threat to democracy? While some politicians and media outlets present it as dangerous to the US, Europe, and Latin America, others hail it as the fix for broken democracies (Moffitt, 2016). Further, what is populism's relationship to the substantive democratic value of a political programme? We appreciate that issues about populism's connection to democracy are not straightforward but need to be considered under specific political practice and actions. However, this study contends that 'post-truth' in politics is one of the drivers of populism and a threat to democracy (Iosifidis and Andrews, 2019).

POLICY ISSUES AND SOCIAL MEDIA

Recently, the social media has facilitated interactive communications between the political elites and the public. Iosifidis and Wheeler (2018) considered how politicians have employed the social media to affect major changes in recent US Presidential campaigns and the European Union (EU) Referendum. In particular, in

the 2016 UK Referendum, social media networks became a vehicle for contested political arguments and 'post-truth' positions defined mainly the Leave camp. For example, it was claimed that the UK Independence Party former leader Nigel Farage's anti-migrant tweets influenced many voters. In the 2016 US Presidential election, the victorious celebrity property tycoon Donald Trump maintained a controversial online presence. He posted tweets about his campaign and engaged in a blatantly hateful online discourse aimed at his political opponents. Such a usage of the social media does not aid democratic representation, but instead it contributes to a greater destabilisation of modern politics. Therefore, the spread of disinformation has to do both with technological processes but also motivated political actors.

Despite the severity of the aforementioned allegations of information warfare occurring on such a significant and instantaneous network, little research work has been conducted with regards to overseeing social media and investigating the ways in which they facilitate populism narratives and the spread of fake news. As Freedman (2018) noted, policy silences made it easier for the rise of powerful and yet unaccountable digital intermediaries through whose channels travels the fake news. He argued that unregulated digital platforms, the pursuit of media coverage and the communication of rage are core to the growth of reactionary populisms and called for a new policy paradigm that is based not merely around ideas of freedom, access and accountability, but on the redistribution that is required to tackle the abuse of media power by large corporations.

However, several problems arise when it comes to regulating social media. Restricting political speech is ultimately a violation of freedom of speech even if it is false. The main reason is because the state, ironically, will then have the power to use and decide what is true and what is false for its own political end (Timer, 2017). As

Feldman (2016) noted, “in the free marketplace of ideas, true ideas are supposed to compete with false ones until the truth wins”. Facebook, as all online sites, is protected under The US Constitution’s First Amendment and freedom of speech principles more generally albeit calls for amendments more suited for fake news and digital communication (for an overview of these issues and a critique of First Amendment theory see Napoli, 2018). Furthermore, through Section 230 of the Communications Decency Act, all online sites have immunity when it comes to political discourse (Timer, 2017).

As a result, the way in which Facebook regulates itself remains an important factor in regard to efforts that stifle online disinformation. For a long time, the company had been criticized for not doing enough (Tufekci, 2016), yet following the 2016 US presidential election it has been noticeably more active (Manjoo, 2017).

ONLINE DISINFORMATION

Disinformation is defined as “false, incomplete or misleading information that is passed, fed, or confirmed to a targeted individual, group, or country” (Shultz and Godson, 1984: 41). Jowett and O’Donnell (2012) define disinformation as black propaganda because of its covert nature and use of false information. The authors connect the term to what was once a KGB division known as *dezinformatsia,* devoted to black propaganda (23-24). They further emphasize only a few years ago disinformation spread across US print media, “to weaken adversaries and [were] planted in newspapers by journalists who are actually secret agents of a foreign country” (24).

Recent studies on disinformation portray a similar picture. Bennett and Livingston (2018: 124) define disinformation as “intentional falsehoods spread as

new stories or simulated documentary formats to advance political goals". Humprecht's definition (2018: 2) adds a profit motive describing disinformation as information that is intentionally created and uploaded on various websites, and thereafter disseminated via social media either for profit or for social influence. The UK's *Disinformation and 'fake news' Report* (DCMSC of the House of Commons, 2019: 7) similarly defines disinformation as, "the deliberate creation and sharing of false and / or manipulated information that is intended to deceive and mislead audiences, either for the purposes of causing harm, or for political, personal or financial gain" and is in line with its European counterpart (European Commission, 2018). These definitions reflect the work of MIT political economists Benkler, Faris and Roberts (2018) who in a largescale study in the US have acknowledged five parties that circulate online disinformation. These are:

1. Bodies close to Russian government
2. Right-wing groups
3. Groups that make money such as those based in Macedonia
4. Formal campaigns using marketing tools (i.e. *Cambridge Analytica*)
5. Peer-to-peer distribution networks

Despite efforts to fight back against online disinformation its diffusion continues to progress as social media platforms expand their user base and deviant initiators adapt to online changes. Deepfakes, for example, a technique for human image synthesis based on artificial intelligence, can create audiovisual and audio content identical to real people. One can imagine during times of unrest how deepfakes could swing opinions one way or another.

The facilitating factor of today's online usage can be traced back to the rubric of 'digital capitalism' (Schiller, 1999). Users connect to online market systems designed to maintain user attention that results in an ominous rise in online behavioral addiction (see Wu, 2016 and Alter, 2017). As an increasing number of users spend more time on social media, the likelihood of disinformation getting shared also rises. Here, the content is considered misinformation since the senders do not know the original story is fake (DCMSC of the House of Commons, 2019). Deviant disinformation agents intentionally make these stories more engaging via emotional appeal, making users to more willingly share it (D'Ancona, 2017).

Disinformation is endemic to digital networks. Carlson (2018) postulates that digital technologies accentuate societal deviancies such as disinformation and that technology itself develops into the main culprit against an existing moral order. Indeed, Jowett and O'Donnell (2012: 159) note, "the very 'democracy' and accessibility of the World Wide Web has made it the most potent force for the spreading of disinformation yet devised". Facebook, primarily a technology company, becomes the possessor of social deviance online. As this happens news media are conversely converted into those institutions upholding moral order. As online disinformation continues to generate moral panics news media paradoxically reclaim legitimacy as the institutions best suited to uphold contemporary public spheres (Boyd-Barrett, 2019). Yet ironically, news media are heavily reliant on social media platforms. Most of them have constructed multiplatform options that includes forming a social media presence (Ju, Jeong and Chyi, 2014; Hagvar, 2019). Revenues raised from news media websites are linked to eyes on screen; therefore, all major publishers create Facebook pages in an effort to drive traffic from the social media platform to their own websites. The result is Facebook's growing influence in news

consumption. According to the *New York Times,* Facebook 'has become the largest and most influential entity in the news business, commanding an audience greater than that of any American or European television news network, any newspaper or magazine in the Western world and any online news outlet' (Manjoo, 2017). Until news media publishers find other revenue sources their efforts to drive traffic from social media will continue as will the potential for disinformation and misinformation to spread through deviants portraying themselves as legitimate news media (Hagvar, 2019). News media's use of social media also adds to the obfuscation of content misinterpreted as fake. In Tandor, Lim and Ling's (2018: 141) typology of fake news, they include news satire and parody. Jaster and Lanius (2018: 214-15) add journalistic errors and highly selective reporting to the list. We can further add opinion pieces that social media machine learning might identify as bad content and therefore make it easier for deviants to cloak their own content under these categories. By broadening the context of online disinformation across an ever-expanding news ecosystem (Picard, 2014) the presence of digital news media on social media platforms might in fact contribute to the threat of online disinformation.

FACEBOOK, STAKEHOLDER RELATIONS AND DISINFORMATION

As illustrated, Facebook's impact in contemporary life is unequivocal. Launched only 15 years ago, at the time of writing, it reaches over two billion people every month, while 1.2 billion users visit the platform daily. In other words, Facebook is currently the internet's most visited website both in terms of viewed pages and time spent. Because of its influence, and as we have portrayed above, Facebook has come under scrutiny for its potential to be used to spread inappropriate content rapidly and globally. The platform has been used to promote violent acts of

terrorism, distort presidential election results, influence perceptions and sway public opinion during important democratic moments. The claim that Pope Francis had endorsed Mr. Trump for president, although fake, was shared nearly a million times (Isaac, 2016).

A large number of fake Facebook pages and accounts acting as sources of disinformation are suspected of having ties with Russian and Iranian entities and come with specific agendas therefore creating a form of 'information warfare' through Facebook. In the UK the government has accused Russia of meddling in elections by spreading disinformation (DCMSC of the House of Commons, 2019). According to the *Washington Post* (Priest, Jacoby and Bourg, 2018), Facebook was warned by several activists, civil society organisations and journalists from around the world about such cases but the tech giant was palpably slow to act. In fact, several former employees of Facebook admitted they knew about these cases but could do little about it (DCMSC of the House of Commons, 2019).

Facebook's privacy tribulations began as early as 2006 with its introduction of newsfeed (Tufekci, 2018), but it was not till it became a publicly traded company on 18 May 2012 that corporate pressure mounted. That same day, the one-time startup built and designed by a twenty-year-old Mark Zuckerberg in his Harvard dorm room less than ten years earlier, became the world's largest valuated company at $US104 billion. Tacit and explicit investor pressure was immediate as its share price plummeted the moment trading commenced. The sudden drop in its share price resulted in several lawsuits and a severe hit to the company's reputation. 57 per cent of the shares sold at the time were from Facebook insiders, while GM notably withdrew 10 million dollars of its advertising budget due to its lack of confidence in how effective Facebook's advertising services were (Walton, 2018). With investor

pressure growing, Facebook needed to quickly find ways to increase revenues. As a consequence, the very same year it went public it pivoted from an online payment for games and applications to an advertising-driven business model delivered mainly on smart phones (DCMSC of the House of Commons, 2019: 26). In order to achieve this Facebook began to harvest personal information from its users that could in turn be used to apply more targeted advertising messages. User information was allowed to be used by app developers (for a price), that in turn led to the 2018 *Cambridge Analytica* data scandal (Rosenberg, Confessore and Cadwalladr, 2018). As described in the DCMSC report (2019: 41), Ashkan Soltani, former Chief Technologist to the US Federal Trade Commission, gave a damning report regarding his definition of Facebook: "it is either free – there is an exchange of information that is non-monetary – or it is an exchange of personal information that is given to the platform, mined, and then resold to or reused by third-party developers to develop apps, or resold to advertisers with advertisers".

In spite of the damage to its reputation, particularly in the wake of the well-documented 2018 *Cambridge Analytica* incident, the organisation has continued to prosper. It recently announced better than expected results in earnings and revenue as well as in its continued growth in users (Wong, 2019). In fact, the company has steadily grown every year since it went public. At the time of writing Facebook has a market capitalization of $US477 billion (YahooFinance, 2019). With such influence, the company has entered a phase that requires it to be more transparent but also more responsible towards its investors and stakeholders. As a *Fortune 100* company Facebook finds itself constantly having to manage its reputation and investor expectations (see Nicoli and Papadopoulou, 2017), while still being seen as the company 'that connects the world'. This is a difficult feat for any public company let

alone a social media platform used by almost half the world's population. Exacerbating the scrutiny is the rise in demand for innovation journalism and technology news (Gynnild, 2013), whereby audiences seek to consume news relating to technology companies and disruptive digital transformations.

As a publicly traded company Facebook adheres to several Securities and Exchange Commission regulations regarding how it discloses information. The company is expected to keep, among several credentials, archives of annual reports, quarterly earnings, valuations and communication that are sent out to news media often in the form of its own news releases. Most public companies do so transparently, keeping archives on their websites often in the form of a dedicated newsroom that stores all announcements. Facebook is no exception. As a Silicon Valley company with huge cultural significance, it might even be expected to innovate in how it communicates to its stakeholders and certainly in how it contributes to society (Nicoli and Komodromos, 2019). As such, the company's newsroom (newsroom.fb.com) is considerably rich in content. It has eight tabs (home, news, company info, directory, media gallery, inside feed, public policy and investor relations), a search directory and an email for inquiries (press@fb.com).

METHODOLOGY

The research question of this study is:

*RQ***:** What are Facebook's announcements on tackling online disinformation, misinformation, false news and fake news?

Research Design

In order to assess the research question two qualitative umbrella approaches are considered, discourse analysis and content analysis. The strength of discourse

analysis lies in how it comprehends the way social power is abused and inequality is enacted (see for example Van Dijk, 1997). The constructivist nature of discourse analysis highlights a subjective understanding of the analysed data in which assumptions are not only made, they are encouraged. According to Schreier (2012: 45), "one of the basic assumptions underlying discourse analysis is that language does not represent reality, but that it contributes to the construction of reality, and to the construction of social reality in particular". On the other hand, the origins of content analysis lie in Harold Lasswell's classic communication process of who says what to whom and with what effect (Bloor and Wood, 2006). The method offers a *realist perspective* whereby *assumptions* about reality are avoided and therefore data is seen more for what it sets out to be from its encoder. Qualitative content analysis is used, in this study too, as a systematic coding and categorization approach for exploring and interpreting large amounts of textual data to determine, in an *unobtrusive* manner, patterns of words used, their frequency, their relationships and their structures of communication (Vaismoradi *et al*., 2013: 400).

Data Collection

The collected data draws on the period 16 November 2016 and 4 March 2019. The following words are examined separately in four different searches across the aforementioned time period: *disinformation, misinformation, false news* and *fake news.* The four searches yielded 108 results that were categorized as 108 units of analysis (UOA 1-108). A sample of the results of the search are illustrated in Appendix A.

The results include a Zuckerberg post (UOA 1), a Facebook whitepaper (UOA 13), a large-scale interview of Mark Zuckerberg by well-known journalist and blogger Ezra Klein in vox.com (UOA 33), and an interview with Chris Cox in Wired

Magazine (UOA 57). Zuckerberg's Facebook post is a 16 November 2016 pronouncement on Facebook and misinformation, while the white paper titled, 'Information Operations and Facebook' covers, in an overarching manner, the way in which Facebooks handles bad content. Several of the Facebook announcements are responses to news media criticism in which we consider in our analyses. Others are short text with links to further sources of information or in-house videos that were all investigated and considered in the analyses. The time period chosen for analysis begins with Zuckerberg's 2016 post since it is a seminal announcement a week following Donald Trump's win in the 2016 presidential elections.

Data Analysis

The Coding Frame: The Three Dimensions

In order to structure the study to address the research question three dimensions (main categories) are applied to classify and interpret the 108 units of analysis. These are: *proactive announcements directly dealing with tackling online disinformation, reactive announcements and discussion pieces* and *residual announcements* (miscellaneous announcements that do not fit in the other two, see Schreier, 2012). As illustrated, one of the key criticisms of Facebook has been that it does not respond soon enough (if at all) to accounts of disinformation occurring on its platforms. Indeed, in front of congress Zuckerberg noted that "Facebook needs to take a more proactive role in issues such as fake news" (CNET, 2018). If Facebook is taking proactive steps in combating disinformation, then a self-reforming approach can be warranted, and we might easier conclude that Facebook is taking steps to combat disinformation.

Subcategories for the analysis are chosen in a data-driven (inductively) manner whereby categories emerge from the analyzed content (see Schreier 2012:

60). A concept-driven (deductively) approach is avoided since theories and approaches in combating disinformation are still unclearly defined. Had a concept-driven approach been used based on our own understanding of how disinformation can be combated, subcategories might have incorrectly been placed or missed altogether. Similarly, in regards the reactive announcements, we could not anticipate what, how and by whom Facebook would be reacting to, therefore its dimension's sub-categories are also created solely on a data-driven inductive approach emerging from the content. Appendix B illustrates Dimension 1 and its ten subcategories (D1.1-D1.10), Appendix C, Dimension 2 with eight subcategories (D2.1-D2.8) and Appendix D, Dimension 3, seven subcategories (D3.1-D3.7). As illustrated in Chart 1., the majority of the content analysed is categorized and coded in D1 showing the proactive manner in which Facebook is attempting to combat disinformation. D2 and D3 yielded less results that were considered in our analyses far less significant than D1.

Please insert Figure 1 here

Findings and Sub-Categories

The following subcategories emerge from the three dimensions we designed in our content analyses.

Dimension 1 (Appendix B): Proactive Announcements Directly Dealing with Tackling Disinformation (D1): Total of 79 Announcements

D1.1: Generic - All articles that fall into this category address a wider picture of what Facebook is announcing in terms of proactively tackling disinformation on its platforms. These articles list a series of Facebook actions without going into much

detail regarding each one. Zuckerberg's own post in 2016 (UOA 1) does precisely this whereby he lists the areas Facebook is working on in how it tackles disinformation. Overall, this was the largest of the subcategories with 32 of 79 dimension one announcements.

D1.2: Penalizing Content (In Newsfeed) – Facebook does not remove all suspicious and bad content; it rather feels enough is done if such content is penalized to a lower ranking on its newsfeed. Thousands of pieces of information in the form of Facebook posts are displayed on users' timelines yet on average only around 300 are scrolled and viewed. Facebook identifies several bad practices that it penalizes to a lower rank in its users' newsfeeds. These are, clickbaits, ad farms, sensationalism pieces and misinformation. It also chooses to block Facebook pages from been seen from advertising if they are sharing stories detected as false either by other users or third-party checking organisations. D1.2 had four dedicated announcements of the 79 in dimension one.

D1.3: Related Articles Quality – On 20 December 2017 Facebook announced it would be using related articles instead of disputed flags for posts that have been detected as false. As the company notes, it is to help users receive more context about a story by putting related articles next to a false news story. This according to Facebook leads to less shares than disputed flags and therefore less misinformation. Flagging a post actually leads to more shares so related articles is a more appropriate way of dealing with the disputed article without having to take it down. Our search (using the four keywords) yielded one dedicated announcement albeit it should be noted that in a

separate search on ‘related articles’ further announcements on the topic were displayed.

D1.4: Disrupting Economics of Disinformation Operations – This subcategory was identified in Facebook’s generic announcements on disinformation (D1.1). In them it has shown it is combating organizations that attempt to generate money by posting bad content. This is done by detecting through third-party checking organizations, advertising policies and by applying machine learning. Our analyses did not yield any dedicated announcements on the topic.

D1.5: Building New Products – Although there is a palpable overlap with other subcategories (e.g. ‘related articles’ is a new product), these were announcements we felt were worthy of a category. New products include mechanisms of transparency of (political) advertising, encouraging local news publishers, detection products (D1.9), and new security protection measures for political campaigns (see also 1.10). The ads transparency tool allows users to visit Facebook pages and to see what ads the advertiser is running and whether or not they are shown to the user. D1.5 had four dedicated announcements of the 79 in dimension one.

D1.6: Facebook Verification – Facebook has created a page verification (blue flag) procedures that offers further accountability regarding the publishers of Facebook pages. In parallel to this the company has improved the verification of naming (users, groups and pages) and therefore minimizing fake names. This subcategory yielded two dedicated announcements.

D1.7: Removing Content and Offenders – Facebook has identified certain violations that sanction the company from removing the content. These include cloaking, repeatedly sharing fake news from detecting mechanisms, taking down fake accounts such as those linked with the Internet Research Agency (Kremlin linked troll group) often with the help of tip offs from stakeholders (e.g. third-party fact checkers) or government institutions (e.g. FBI), coordinated inauthentic behaviour and violations of its community standards (violence and criminal behaviour, objectionable content, respect of intellectual property etc.). Cloaking is when violators disguise the true destination of an advertisement or post in order to bypass the review processes implemented by Facebook. Coordinated inauthentic behaviour is when networks of accounts or pages corroborate to mislead who they are and what they are doing. D1.7 yielded 19 dedicated announcements.

D1.8: Reaching Out to Stakeholders and Establishing Partnerships (Excluding Third-Party Fact Checkers) – Facebook has set up partnerships and projects with various stakeholders and institutions regarding how it combats disinformation. These include news media (e.g. *the Facebook Journalism Project* promoting news literacy and training), government authorities, other technology companies via a procedure known as threat exchange, and a partnership with the Think Tank *Atlantic Council* in order to assist in identifying real-time insights on emerging threats. D1.8 yielded one dedicated announcement.

D1.9: Detection – This subcategory involves how Facebook detects disinformation. The manner in which it detects disinformation is by machine learning classifiers (feeding a computer with examples of bad content to find itself patterns), using a

number of independent third-party fact-checkers from around the world, easy reporting mechanisms designed for users, artificial intelligence and computer vision technologies, partnering with election integrity teams on an ad hoc basis during elections around the world (see also D1.10) and using trustworthy surveys. Our analyses yielded 10 dedicated announcements within this subcategory.

D1.10: Election Tools – Facebook has taken various measures to protect democratic processes from significant elections from around the world using technologies and teams. Technologies include issues tabs on candidate pages, candidate info tools, vote planning tools, ad hoc fact-checkers and war rooms set up with Facebook teams from across various departments (e.g. war rooms were set up for the Brazil and US midterm elections with specialists from numerous departments such as security and communication). Overall, five announcements were found that fell into this subcategory.

Please insert Figure 2 here.

Dimension 2 (Appendix C): Reactive Announcements and Discussion (D2): Total of 18 Announcements

Announcements coded into this dimension do not declare actual approaches regarding how Facebook fights back against disinformation. These announcements rather prompt discussion on disinformation or respond to previous announcements or criticisms. Furthermore, much of the content from D2 does not prioritize disinformation but is rather a part of another topic. Chart 3. illustrates the breakdown. Almost half the announcements involve a series of discussions with academics,

intellectuals and Facebook employees on the role of social media and democracy; within these announcements the topic of disinformation is deconstructed under a larger picture of social media's role in democracy. The second most popular announcements here involve the way in which Facebook responded to articles published in popular news media. UOA 86 for example disputes several 'inaccuracies' of a *New York Times* article published on 14 November 2018 addressing issues of ongoing Russian investigations, Facebook's connection to a public affairs agency (Definers), whether President Trump's comments broke community standards, Facebook's commitment to fighting fake news and other issues.

Please insert Figure 3 here.

Dimension 3 (Appendix D): Residual Announcements (D3): Total of 11 Announcements

Dimension 3 yielded 11 results from the 108 units of analyses (see chart 4). These were broken down into seven subcategories that consisted of announcements in which the four keywords were not at all a priority of the announcement.

Please insert Figure 4 here.

DISCUSSION FOR FUTURE RESEARCH

Facebook endevours to combat disinformation by tweaking technologies and policies of how users spend time on the platform. Detection and categorization of bad content are Facebook's most common approaches to achieving this; both require machine learning and AI to assist human moderation, especially in regard to photos

and video fact-checking. Facebook has not announced in any detail the connection between human moderation and machine learning AI approaches to identifying disinformation. In listing the weaknesses of AI and machine learning, Scharre notes "if the data don't represent the system's operating environment well, the system can fail in the real world…AI systems can go from supersmart to superdumb in an instant" (2019). Areas of study where AI systems of detection need to improve are in how to treat satire, opinion pieces, deepfakes and cloaking since these seem to be common ways deviant initiators will attempt to share fake news on Facebook.

Throughout the analysis Facebook identified itself as not being an arbiter of truth, no doubt having the First Amendment and Section 230 of the Communications Decency Act in mind. By doing so the company is showing its reluctance to remove content from its platforms even when it knows it is bad. It rather sets community standards and policies sanctioning itself to categorize content and penalize it by lowering its rank. This too is one area which requires more research since setting specific guidelines from regulators (or amending the First Amendment) might help social media platforms remove bad content altogether. The caveat here is that a balance is required in regulating content without threatening universal freedoms of expression. Finally, more exploration is needed in disrupting the financial flows of disinformation operations but also of coordinated inauthentic behaviour for ideological purposes. This could be through myth-busting, applying international pressure and strategic communication against deviant actors.

Conclusion

Social media are rife with opportunities for those who seek to destabilize societies through disinformation either for ideological purposes, financial gains or

both, particularly before elections. This occurs even in mature democracies as we have illustrated the connection between populist movements in liberal democracies triggered by deviant actors sharing disinformation on social media platforms. Regardless of whether they are to blame, social media are certainly used as accelerants for swaying public perceptions in directions that meet specific agendas or divide people through tribal discourse where users' biases are confirmed through selective exposure.

Online disinformation needs to be identified as a multifaceted problem; one that requires multiple approaches to resolve. Governments, regulators, think tanks, the academy and technology providers need to join forces to better shape the next internet with as less online disinformation as possible. While some level of self-regulation is applied at Facebook there seems to be too much reliance on AI systems to detect bad content. If this is countered with bad content that is also created by AI systems then it might not suffice to counter disinformation particularly when it comes in the form of satire, opinion pieces and deepfakes. In such cases one wonders whether human moderation (e.g. fact-checkers) is enough on a platform as large as Facebook. Also, might the First Amendment and Section 230 of the Communications Decency Act be amended in order to allow social media to take more actions on dubious content in a shorter amount of time?

REFERENCES

Alter A (2017) *Irresistible: The Rise of Addictive Technology and the Business of Keeping us Hooked*, New York, Penguin

Benkler Y, Farris R and Roberts H (2018) *Network Propaganda*. NY: Oxford University Press.

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

Boyd-Barrett O (2019) Fake news and 'RussiaGate' discourses: Propaganda in the post-truth era, *Journalism* 20(1): 87-91.

Bloor M and Wood F (2006), *Keywords in Qualitative Methods: A Vocabulary of Research Concepts* (1st ed). London: Sage.

Cairncross F (2019) *The Cairncross Review: A Sustainable Future for Journalism.* 12 Feb. [Available at]: <https://www.gov.uk/government/publications/the-cairncross-review-a-sustainable-future-for-journalism> (accessed 15 March 2019).

Carlson M (2018) Fake news as an informational moral panic: the symbolic deviancy of social media during the 2016 US presidential election, *Information, Communication & Society*, DOI: 10.1080/1369118X.2018.1505934.

Choy Y (2018) Online political public relations as a place-based relational practice: A cultural discourse perspective. *Public Relations Review* 44(5): 752-761.

CNET (2018) *Zuckerberg Senate Hearing Highlights in 10 minutes*, YouTube video, added by CNET [Online]. [Available at]: <https://www.youtube.com/watch?v=EgI_KAkSyCw> [Accessed 3 March 2019].

DCMSC (2019) *House of Commons Digital, Culture, Media and Sport Committee, Disinformation and 'fake news': Final Report, Eighth Report of Session 2017–19.* [Available at]: <https://www.parliament.uk/business/committees/committees-a-z/commons-select/digital-culture-media-and-sport-committee/news/fake-news-report-published-17-19/> (accessed 28 March 2019).

D'Ancona M (2017) *Post Truth: The new war on truth and how to fight back*. London: Ebury Press.

European Commission (2018) *'A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation'*, [Available at]: <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation> (accessed 3 September 2018).

Feldman N (2016) Fake News May Not Be Protected Speech. *Bloomberg View.* 23 Nov. [Available at]: <https://www.bloomberg.com/view/articles/2016-11 -23/fake-news-may-not-be-protected-speech> (accessed 28 October 2018).

Fitzi G, Mackert J and Bryan ST (2018) *Populism and the Crisis of Democracy: Volume 1: Concepts and Theory*. London: Routledge.

Flew T (2017) The 'Theory' in Media Theory: The 'Media-Centrism' Debate. *Media Theory* 1(1): 43-56.

Freedman D (2018) Populism and media policy failure. *European Journal of Communication* 33(2): 122-139.

Fukuyama F (2018) Against Identity Politics: The New Tribalism and the Crisis of Democracy. *Foreign Affairs* 97(5): 90 – 114.

Gynnild A (2013) Journalism innovation leads to innovation journalism: The impact of computational exploration on changing mindsets. *Journalism* 15(6): 713-730.

Hågvar B (2019) News Media's Rhetoric on Facebook. *Journalism Practice*. DOI: 10.1080/17512786.2019.1577163.

Hinde S and Dixon J (2007) Reinstating Pierre Bourdieu's contribution to cultural economy theorizing. *Journal of Sociology*. 43(4): 401 – 420.

Humprecht E (2018) Where 'fake news' flourishes: a comparison across four Western democracies. *Information, Communication & Society*, 21: 1-16.

Inglehart RF and Norris P (2016) *Trump, Brexit, and the Rise of Populism: Economic Have-Nots and Cultural Backlash*. Faculty Research Working Group Series. Harvard Kennedy School.

Iosifidis P and Wheeler M (2018) Modern political communication and Web 2.0 in representative democracies. *Javnost/The Public* 25(1-2). DOI: 10.1080/13183222.2018.1418962.

Iosifidis P and Andrews L (2019) Regulating the Internet Intermediaries in the Post-Truth World: Beyond Media Policy?' *International Communication Gazette*. Accepted version: http://openaccess.city.ac.uk/20517/.

Isaac M (2016) *Facebook* Mounts Effort to Limit Tide of Fake News. *The New York Times*. 15 Dec. [Available at]: <https://www.nytimes.com/2016/12/15/technology/*Facebook*-fake-news.html7_F0.> (accessed 3 September 2018).

Isaac M & Roose K (2018) Disinformation Spreads on WhatsApp Ahead of Brazilian Election. *The New York Times*. 19 Oct. [Available at]: <https://www.nytimes.com/2018/10/19/technology/whatsapp-brazil-presidential-election.html?rref=collection%2Fbyline%2Fmike-

isaac&action=click&contentCollection=undefined&region=stream&module=stream_unit&version=latest&contentPlacement=2&pgtype=collection> (accessed 19 October 2018).

Jaster R and Lanius D (2018) What is fake news? *Versus* 2(127): 207 – 227.

Jowett H and O'Donnell V (2012), *Propaganda and Persuasion*, 5th Edition, London: Sage.

Ju A, Jeong S and Chyi H (2014) Will Social Media Save Newspapers? *Journalism Practice* 8(1): 1-17. DOI: 10.1080/17512786.2013.794022.

Manjoo F (2017) Can Facebook Fix its Own Worst Bug? *The New York Times Magazine*. [Available at]: <https://www.nytimes.com/2017/04/25/magazine/can-Facebook-fix-its-own-worst-bug.html?_r=0> (accessed 3 September 2018).

Mouffe C (2018) *For a Left Populism*, New York: Verso.

Mudde C (2007) *Populist Radical Right Practices in Europe*. NY: Cambridge University Press.

Moffitt B (2016) *The Global Rise of Populism: Performance, Political Style, and Representation*. USA: Stanford University Press.

Napoli PM (2018) What If More Speech Is No Longer the Solution? First Amendment Theory Meets Fake News and the Filter Bubble, *Federal Communications Law Journal*, 70(1).

Nicoli N and Komodromos M (2019) CSR Communication in the Digital Age: The Case of The Bank of Cyprus. *Cases on Corporate Social Responsibility and Contemporary Issues in Organizations*, Antonaras A and Dekoulou E (Eds.). Pennsylvania IGI Global. 71 – 89.

Nicoli N and Papadopoulou E (2017) Building and Protecting Reputation Through Trip Advisor: A Case Study for the Cyprus Hotel Industry. *EUROMED Journal of Business*, 12(3): 316 – 334.

Norris P and Inglehart RF (2019) *Cultural Backlash: Trump, Brexit, and Authoritarian Populism*, Cambridge: Cambridge University Press

Picard R (2014) Twilight or New Dawn of Journalism? Evidence from the Changing News Ecosystem. *Journalism Studies* 15(4): 1-11.

Priest D Jacoby J & Bourg A (2018) Russian disinformation on Facebook targeted Ukraine well before the 2016 U.S. election. *The Washington Post.* 29 Oct. [Available at]: <https://www.washingtonpost.com/business/economy/russian-disinformation-on-Facebook-targeted-ukraine-well-before-the-2016-us-election/2018/10/28/cc38079a-d8aa-11e8-a10f-b51546b10756_story.html?noredirect=on&utm_term=.f5602ecef56b>, (accessed 29 October 2018).

Rosenberg M Confessore N and Cadwalladr C (2018) How Trump Consultants Exploited the Facebook Data of Millions. *The New York Times*. [Available at]: <https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html> (accessed 19 February 2019).

Scharre P (2019) Killer Apps: The Real Danger of an AI Arms Race, *Foreign Affairs*, May/June.

Schiller D (1999) *Digital Capitalism: Networking the Global Market System*, Massachusetts, MIT Press.

Schreier M (2012) *Qualitative Content Analysis in Practice*. London: Sage.

Shultz R and Godson R (1984) *Dezinformatsia: Active Measures in Soviet Strategy*: Oxford, Pergamon-Brassey.

Tandor E, Lim Z and Ling R (2018) Defining 'Fake News', *Digital Journalism* 6(2) 137-153.

Timmer J (2017) Fighting Falsity: Fake News, Facebook, and the First Amendment. *Cardozo Arts & Entertainment* 35(3): 669–705.

Tufekci Z (2016) Mark Zuckerberg in Denial. *The New York Times: Opinion*. [Available at]: <https://www.nytimes.com/2016/11/15/opinion/mark-zuckerberg-is-in-denial.html?mcubz=0&_r=0> (accessed 23 September 2018).

Tufekci Z (2018) Why Zuckerberg's 14-Year Apology Tour Hasn't Fixed Facebook [Available at]: < https://www.wired.com/story/why-zuckerberg-15-year-apology-tour-hasnt-fixed-facebook/> (accessed 19 April 2019).

Van Dijk T (1997) Editorial: Applied Discourse Analysis. *Discourse and Society*. 8(4) 451 – 452.

Vaismoradi M Turunen H and Bondas T (2013) Content Analysis and thematic analysis: Implications for conducting a qualitative descriptive study. *Nursing & Health Sciences*. 15: 398-405.

Walton J (2018) When did Facebook go public? [Available at]: <https://www.investopedia.com/ask/answers/111015/when-did-facebook-go-public.asp> (accessed 23 September 2018).

Wong J (2019) Facebook posts record profits despite year of scandals. [Available at]: <https://www.theguardian.com/technology/2019/jan/30/facebook-fourth-quarter-profits-revenues-earnings> (accessed 23 February 2019).

Wu T (2017) *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*, New York, Alfred A. Knopf Press.

Yahoo Finance (2019) *Facebook Inc. Summary*. [Available at]: <https://finance.yahoo.com/quote/FB/?guccounter=1> (accessed 23 February 2019).

Zuckerberg M (2019) Mark Zuckerberg. The Internet needs new rules: Let's start with these four areas. [Available at]: <https://www.washingtonpost.com/opinions/mark-zuckerberg-the-internet-needs-new-rules-lets-start-in-these-four-areas/2019/03/29/9e6f0504-521a-11e9-a3f7-78b7525a8d5f_story.html?utm_term=.842f5cb21e4e> (accessed 1 April 2019).

APPENDIX A (SAMPLE OF UNITS OF ANALYSIS UOA 1-7)

UOA	TITLE / UNIT OF ANALYSIS	HEADLINE	DATE	SEARCH KEYWORD	SOURCE
1	Zuckerberg post	"A lof of you have asked what we're doing about misinformation"	16/11/2016	on post truth	Mark Zuckerberg FB post
2	Addressing Hoaxes and Fake News	We're committed to doing our part to address the issue of fake news and hoaxes on Facebook.	15/12/2016	**Fake news**	NEWSROOM
3	Continuing Our Updates to Trending	We're announcing three updates to Trending, a feature that shows people popular topics being discussed on Facebook that they might not see in their News Feed.	25/01/2017	**Fake news**	NEWSROOM
4	A New Educational Tool Against Misinformation	False news and hoaxes are harmful to our community and make the world less informed.	06/04/2017	**Misinformation / False News**	NEWSROOM

5	Working to Stop Misinformation and False News	We know people want to see accurate information on Facebook – and so do we.	06/04/2017	**Misinformation / False News**	NEWSROOM
6	Reducing Links to Low-Quality Web Page Experiences	We hear from our community that they're disappointed when they click on a link that leads to a web page containing little substantive content and that is covered in disruptive, shocking or malicious ads.	10/05/2017	**Misinformation / False News**	NEWSROOM
7	Verified Pages and Profiles	https://www.facebook.com/animalyouth?fref=ts	29/05/2017	**Fake news**	NEWSROOM

APPENDIX B: DIMENSION 1 - Proactive announcements directly dealing with tackling disinformation			
SUB-CATEGORY TITLE	**CODE**	**UNITS OF ANALYSIS**	**%**
GENERIC	D1.1	1, 2, 4, 5, 6, 9, 14, 16, 29, 30, 31, 34, 35, 38,42, 44, 45, 46, 47, 48, 49, 50, 58, 65, 79, 85, 97, 98, 99, 102, 104, 107	41%
PENALIZING CONTENT (IN NEWSFEED)	D1.2	7, 10, 11, 27	5%
RELATED ARTICLES QUALITY	D1.3	17	1%
DISRUPTING ECONOMICS OF DISINFORMATION OPERATIONS	D1.4		0%
BUILDING NEW PRODUCTS	D1.5	32, 51, 57, 72, 90	6%
FACEBOOK VERIFICATION	D1.6	8, 19	3%

REMOVING CONTENT AND OFFENDERS	D1.7	12, 13, 56, 61, 62, 66, 67, 75, 80, 82, 84, 87, 91, 94, 95, 100, 101, 103, 105	24%
REACHING OUT TO STAKEHOLDERS & ESTABLISHING PARTNERSHIPS	D1.8	108	1%
DETECTION	D1.9	21, 33, 37,43, 53, 54, 55, 70, 71, 76	13%
ELECTION TRACKING	D1.10	52, 60, 73, 77, 81	6%

APPENDIX C: DIMENSION 2 - Reactive announcements & Discussion pieces

SUB-CATEGORY TITLE	CODE	UNITS OF ANALYSIS	%
RUSSIAN ADS	D2.1	15	6%
SOCIAL MEDIA AND DEMOCRACY	D2.2	22, 23, 24, 25, 26, 36, 83, 93	44%
RESEARCH AT FACEBOOK FOR CREATING NEW PRODUCTS	D2.3	59	6%
ENFORCEMENT EFFORTS AND TRANSPARENCY REPORTING	D2.4	39	6%
FREEDOM OF EXPRESSION AND FACEBOOK	D2.5	63, 64	11%
WOMEN AT FACEBOOK	D2.6	68A/B	6%
RESEARCH ON LESS DISINFORMATION ON FACEBOOK	D2.7	78	6%
RESPONDING TO CRITISISM	D2.8	86A/B, 88, 92A/B/C	17%

APPENDIX D: DIMENSION 3 - Residual Announcements			
SUB-CATEGORY TITLE	**CODE**	**UNITS OF ANALYSIS**	**%**
PERSONAL WELL-BEING AND FACEBOOK USAGE	D3.1	18, 74, 106	27%
PREVENTION OF HARASSMENT AND BULLYING	D3.2	20	9%
COMMUNITY LEADERS	D3.3	28	9%
BAD CONTENT	D3.4	40, 41	18%
SUICIDE AND AI	D3.5	69	9%
HOW FACEBOOK WORKS	D3.6	3, 89	18%
D3.7 WHAT KIND OF INTERNET DO WE WANT	D3.7	96	9%