The Weaponization of Cloud-based Social Media:

Prospects for Legislation and Regulation

Barry Cartwright
School of Criminology
Simon Fraser University
Burnaby, Canada
Email: bcartwri@sfu.ca

George R S Weir
Department of Computer & Information Sciences
University of Strathclyde
Glasgow, Scotland, UK
Email: george.weir@strath.ac.uk

Lutfun Nahar
Department of Gender, Sexuality, and Women's Studies
Simon Fraser University
Burnaby, Canada
Email: lutfun_nahar@sfu.ca

Karmvir Padda
Richard Frank
School of Criminology
Simon Fraser University
Burnaby, Canada
Email: {karmvir_padda; rfrank}@sfu.ca

***Abstract*—The 2016 U.S. Presidential election and the 2016 U.K. Brexit referendum are notable for the contemporaneous efforts by Russian-based trolls to manipulate public opinion through Cloud-based social media. Such disinformation warfare raises serious concerns about the risk of negatively influencing democratic processes, as well as the need for viable defensive measures. Responses are required in terms of technical means to detect and counter such "fake news," as well as legal proscriptions that can serve to control such threatening activities. The present paper addresses this disinformation warfare scenario, describes our current research and technical work in this area, and reviews legal precedents that shed light on the complexities and pitfalls that legislators and regulators encounter when seeking to remediate the threat.**

Keywords-Cloud-based social media; disinformation warfare; "fake news"; legislation.

I. INTRODUCTION

Legislators and government regulatory agencies worldwide face a serious challenge when it comes to the regulation of emerging online threats, such as the type of weaponization of Cloud-based social media that was witnessed in connection with the U.S. Presidential election [1] and the U.K. Brexit referendum [2], [3]. The Internet Research Agency, often referred to as the Russian troll army, deliberately distributed so-called "fake news" stories via social media accounts that had been set up for that express purpose. In the U.S., for example, these "fake news" stories heavily favored Donald Trump over Hillary Clinton in the U.S. Presidential election [2], [4], [5]. According to Special Counsel Robert S. Mueller III's recently released report into Russian interference in the U.S. Presidential election [5], these Facebook and Twitter accounts targeted certain groups, such as Blacks (through the Blacktivist Facebook page), Southern Whites (through the Patriototus Facebook page), and the right-wing anti-immigration movement (through the Secured Borders Facebook page), as well as through Twitter feeds such as @TEN_GOP (which claimed to have a connection to the Republican Party of Tennessee), and @America_1st (an anti-immigration account). In the U.K., the "fake news"—which largely stoked Islamaphobic and anti-immigration passions—made extensive use of Twitter, employing Twitter handles such as ReasonsToLeaveEU, or #voteleave [3], [6], [7].

Social network platforms themselves are coming under increasing pressure from legislators and government regulatory agencies to create and put into action their own in-house policies, practices and procedures for dealing with this issue. To illustrate, Mark Zuckerberg, the CEO of Facebook, was grilled by the U.S. Congress in April 2018 regarding the (witting or unwitting) involvement of Facebook and Instagram in the Russian hostile influence campaign during the run-up to the 2016 U.S. Presidential election [8], [9]. At almost the same time, Mike Schroepfer, the chief technology officer of Facebook, faced a similar hearing in front of a Parliamentary Committee in the U.K. regarding fake accounts, political advertising, and the role of Cambridge Analytica in voter-targeting [10]. In Canada, Robert Sherman, who was the deputy privacy officer for Facebook, and Kevin Chan, who was in charge of Facebook's public policy for Canada, were questioned about the role that Facebook and Cambridge Analytica played in both the U.S. election and the Brexit referendum, and about possible violations of Canadian privacy law [11], [12]. On all three occasions, it was indicated that failure on the part of Facebook and its executives to regulate themselves could result in future government action.

While the term "fake news" is commonly used to describe the content of these Russian-sponsored disinformation campaigns, our textual analysis of 2,500 Facebook items posted by the Internet Research Agency—from January 2015 through December 2017, i.e., during the period leading up to, during, and following the U.S.

Presidential election—indicates that an appreciable number of these stories was actually grounded to one extent or the other in "real news" events that had been reported by mainstream media sources [3]. Presumably, these "real news" stories were selected by the Internet Research Agency so as to enflame passions (or to intimidate or otherwise suppress voter turnout) amongst the targeted groups, and that they were deliberately distributed and re-distributed through automated amplification, with the intention of maximizing the potential audience [2], [13]. Nevertheless, the question remains: "how can Western-style democracies enact legislation against and effectively regulate the expression of personal opinion, or for that matter, the dissemination of what in many cases amounts to something approximating 'real news'"?

In Section II, we proceed by outlining the nature of the "fake news" problem. Section III addresses the problem of identification for "fake news". The challenges facing legislation and regulation are considered in Section IV, while we draw preliminary conclusions in Section V.

II. FRAMING THE PROBLEM

Much has been said in recent years about "fake news" and the "post-truth" era [14]. Indeed, some have erroneously attributed the term "fake news" to U.S. President Donald Trump, who is wont to label anything that runs contrary to his own narrative (especially when it comes from traditional news sources such as *CNN* or *The Washington Post*) as "fake news" [15]. However, propaganda—in the form of fake news and other types of disinformation—has been around for millennia, and has been employed with varying degrees of success by political leaders, military leaders and insurgents throughout history [16], [17]. Indeed, it has been argued that contemporary journalistic norms of balance and objectivity are the end product of a backlash against unabashed use of journalistic propaganda during both World Wars, and the manner in which such propaganda has been put to further use by large corporations [18].

Estimates vary, but it has been stated that 44 percent of the U.S. population gets its news from Facebook, whilst 12 percent gets its news from Twitter [19]. In the U.K., 27 percent of the population gets its news from Facebook, and 14 percent from Twitter [20]. In view of the relatively high percentage of individuals who apparently rely on Cloud-based social media for their news, there is reason for concern with respect to the potential for manipulation of sentiment in this environment. In particular, evidence clearly indicates that the Russians made maximum use of social media bots in their 2016 assaults on the U.S. Presidential election and the U.K. Brexit referendum [1], [2], [21], thereby amplifying the content in order to influence a much wider audience.

In 2017, the Central Intelligence Agency, the Federal Bureau of Investigation and the National Security Agency combined forces to produce an intelligence community assessment of Russian efforts to influence the U.S. Presidential election, concluding that Russia deliberately set out to denigrate and discredit Hillary Clinton whilst promoting the candidacy of Donald Trump, pointing a finger directly at Russia's Internet Research Agency (the Russian troll army), and their use—amongst other attack vehicles—of social media [22]. In February 2018, U.S. Special Counsel Robert Mueller, who was appointed to investigate Russian interference in the U.S. election, obtained a grand jury indictment against Russia's Internet Research Agency (which was bankrolled by Yevgeniy Prigozhin, often referred to as "Putin's chef"), Concord Management and Consulting LLC and Concord Catering (both operated by Yevgeniy Prigozhin), Yevgeniy Prigozhin himself, plus a dozen Russian "trolls" who were employed by the Internet Research Agency. The indictment stated that the accused had "operated social media pages and groups designed to attract U.S. audiences," with the accused falsely claiming that those pages and groups were controlled by American activists, and had used social media platforms such as Facebook, Twitter, YouTube and Instagram to advance divisive issues and create dissension [23].

Similar allegations about Russian interference in the Brexit referendum have surfaced, with as many as 150,000 Twitter bots alleged to have been linked back to Russia [24]. British Prime Minister Theresa May has directly accused Russia of planting fake news stories and seeking to sow discord in Western nations [25]. However, the U.K. government does not appear to have pursued this matter as vigorously as the U.S. government, perhaps because they have been more preoccupied with sorting out the actual ramifications of Brexit.

Evidently, the disinformation attacks by Russia on the U.S. Presidential election and the Brexit referendum were able to achieve results that likely would not have been attainable through more conventional military tactics, such as invading or bombing another country. The disinformation tactics employed by the Russians seemingly succeeded in fragmenting the European Union, testing the strength of the North Atlantic Treaty Organization (NATO), and installing an unabashedly pro-Russian figure in the White House, all without firing a single shot. This could be construed as an all-out assault on Western-style democracy.

III. IDENTIFYING "FAKE NEWS"

The difficulty in detecting hostile disinformation attacks on Cloud-based social media lies in the subtleties between fake news and traditional, "trusted" news. Whereas traditional news has the goal of reporting what happens, albeit sometimes with bias, the purpose of fake news is essentially to insert itself into the same discussion, but to twist the facts in such a way that it incites dissension and distrust. While occasionally relying upon and using the same facts, fake news is thought to focus on nuances that are designed to evoke strong sentiments in the reader. Therefore, the differences between fake news and traditional news may not be so much in the facts or the keywords, which are easier to detect, but rather, in the nuances and sentiment present, both of which are more difficult to detect. It has also been thought that these fake news items are crafted in such a way that they spread six times faster than the truth [26]. Thus, the assumption is that there must be a discernible difference between them.

Our ongoing research involves the analysis of 2,500 “fake news” messages posted on Facebook by Russia’s Internet Research Agency between 2015 and 2017, juxtaposed with 2,500 “real news” items which were derived from 87,157 political news articles from October 2015. The data set of “fake news” posts from the Internet Research Agency was collected and assembled by two professors at Clemson University, Darren Linvill and Patrick Warren [27], and made available by data.world.

The 2,500 “fake news” posts were first read and provisionally analyzed in NVivo, a software tool for qualitative analysis. NVivo facilitates codification and visualization of data, and allows for data queries and automatic provisional coding of the entire dataset. It is anticipated that our ongoing NVivo analysis will lead to the detection of finer nuances and hidden meanings in the data set, which might otherwise not be detected through Posit analysis, or through sentiment analysis (once the matching “real news” data set has been assembled). The Posit toolkit generates frequency data and Part-of-Speech (POS) tagging, producing extensive statistics based on textual content such as social media posts. Posit has proven effective in developing machine learning classification applications [28], [29].

The research team is presently assembling an additional matching set of 2,500 “real news” items from 2015 through 2017, using a set of search terms derived from a careful reading and re-reading of the 2,500 “fake news” items in NVivo, particularly as they pertain to real news events reported in more traditional media sources during that time period. However, the lengthy process involved in assembling a matching “real news” data set did not prove itself amenable to automation. Thus, it was decided that the set of 2,500 “real news” articles from October 2015 would suffice for the purposes of preliminary investigation.

A first round of analysis in NVivo indicated that an appreciable number of the so-called 2,500 “fake news” messages posted on Facebook by Russia’s Internet Research Agency were actually grounded in real news. To illustrate, the second message in the data set, posted under the Facebook name “Patriototus,” referred to the removal of a statue of Confederate General Robert E. Lee in New Orleans. The removal of this statue was reported widely by traditional news sources, including such outlets as *CNN*, *The Washington Post* and the *New York Times*. The second message in the data set, posted under the Facebook name “Blacktivist,” talked about 14-year-old Royce Mann and his slam poem on white privilege and police brutality. The twenty-ninth message, posted under the Facebook name “United Muslims for America,” discussed Kadra Mohamed, the first hijab-wearing policewoman in Minnesota. Again, while the Facebook posts sought to target and agitate certain groups, and were selective in the information they recounted and how they presented it, these events described in Facebook were also reported in traditional, “trusted” news sources.

From the first round of NVivo analysis, a set of search terms (key words and key phrases) was generated, based upon a careful comparison of the Facebook posts to actual events that had been reported in mainstream news sources. Apart from being used to inform ongoing coding in NVivo, and to assist in the assembly of an additional matching set of 2,500 “real news” items, these search terms were matched against the “fake news” data set, to investigate the prevalence of “fake news” items that were in fact grounded in “real news.” In particular, the use of uniquely identifiable, named entities, such as people, places, dates and events indicated that at least 13.5 percent of the so-called “fake news” posts were based to one degree or another on these named entities.

This does not mean that the remainder of the “fake news” posts were entirely fictional. Rather, the posts that did not match these named entities were often vague, or quite short, and contained insufficient information to determine whether they were informed by real news events. A case in point would be the message posted in the Facebook group “Secured Borders,” which asked: “Why there's so many privileges and benefits for refugee kids, but American kids forced to grow up in poverty? That's absolutely unacceptable!!” This could conceivably have been informed by real news events, but it would be a stretch of the imagination to arrive at that conclusion. Nevertheless, it is important to recognize that the term “fake news” is likely a misnomer, which in turn has implications when it comes to the legalities surrounding the suppression of such social media activities.

To secure a source of “real news” data for our comparison with the Facebook “fake news”, we obtained a large set of news posts from webhose.io. This set of 87,157 political “real news” articles, all from October 2015, was derived from a wide variety of Web-based news posts. Sources represented include the *WorldNews (WN) Network*, *Independent Television*, *Philadelphia Daily News*, the *Buffalo News*, the *Press of Atlantic City*, *The Wall Street Journal*, *The Washington Times*, *WCAX News*, Vermont, *KFMB-TV*, *Seattle PI*, *The Boston Herald*, *The Chicago Sun Times*, *The New York Times*, *Fox News* and the *BBC*. Following a process of random selection from the full data source, this “real news” set was reduced to 2,500 data items, found to be derived from a total of 172 news sources.

In order to reduce the original “real news” data to the required 2,500 items, several steps were taken: 1) all news items with duplicate content in the text were removed; 2) the maximum character length of the Facebook “fake news” posts was determined to be 2,006 characters, so all “real news” items with a number of characters greater than 2,006 were removed; 3) the average character length of the Facebook “fake news” posts was found to be 280 characters, whilst the initial average character length for the “real news” data was found to be 1,778 characters; thus, some “real news” items with character lengths greater than 1,000, were expunged from the data set in order to bring the average character length closer to that of the “fake news” posts; and 4) the remaining “real news” data were randomized and a sample of 2,500 items was extracted as the final “real news” set, to serve as a comparator for the 2,500 Facebook “fake news” items. The average character size for the 2,500 “real news” items was 376. A visual inspection of character lengths across the two sets of 2,500 items suggested a similar shaped distribution curve.

Initial comparisons of the “fake news” and “real news” items were conducted using a Posit analysis of their message content. On the basis of a character content analysis, a set of features, including the manual classification of positive or negative for “fake news” was generated for each of the 5,000 data items. Using WEKA [30], and the Random Forest tree-based classifier [31], we achieved a surprisingly high 99.8 percent classification success. While these results are preliminary, and may change when the “fake news” data set is juxtaposed with the second “real news” data set that we are presently assembling, this suggests that we may be able to develop an artificial intelligence tool that can harvest relevant information from social media sources, thereby providing government regulatory agencies with scope for the regulation of the weaponization of Cloud-based social media that was witnessed in connection with the U.S. Presidential election and the Brexit referendum.

IV. Prospects for Legislation and Regulation

While there have been discussions about the potential for government regulation of the dissemination of “fake news” through social media, the issue is far too “new” to have produced any legislation. Therefore, for enlightenment, we must turn to previous efforts to legislate and regulate analogous activities.

The United Nations’ *Universal Declaration on Human Rights* states that “everyone has the right to freedom of opinion and expression,” including the right to “impart information and ideas through any media…regardless of frontiers” [32]. That said, legal positions regarding “acceptable speech” vary widely from country-to-country, and from continent-to-continent [33]. A number of European countries, such as the U.K. and Germany, have enacted (and enforced) laws that are consistent with the European Council’s 2008 Framework Decision on *Combating Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law*, which prohibits expressions that promote hatred or deny crimes of genocide [33], [34], [35]. On the other hand, some European nations, such as Italy, Lithuania and France have grappled with the definition of “hate crime,” and have been more lax when it comes to legal enforcement [36].

Unlike the U.S. and Canada, the U.K. does not have a written *Constitution* or *Charter of Rights and Freedoms* [36]. However, the U.K. attempts to comply with E.U. laws that forbid expressions of racism and xenophobia. A recent example would be the 2018 case of *PF v Mark Meechan*, wherein a Sheriff’s Court in Airdrie, Scotland, fined Meechan £800 for posting a “grossly offensive or threatening” video online, to wit, a video that repeated the phrase “Gas the Jews,” and depicted a dog that had been trained to raise its paw in a Nazi salute [37]. Interestingly, the *Meechan* case generated considerable controversy, with an article in *The Guardian* opining that “giving offence is inevitable and often necessary in a plural society,” and that the judge made an error in conflating offensive material with fomenting hatred [38], and another article in the *American Spectator*, declaring that “free speech is dead in Britain” [39]. As well, a high court decision in the 2011 case of *Abdul v DPP* upheld a lower court conviction of five men who shouted slogans such as “burn in hell,” “baby killer” and “cowards” at a parade of British soldiers, determining that the right to “freedom of expression” under Article 10 of the European Convention on Human Rights did not apply, as the prosecution was “necessary and proportionate” [40].

Although the First Amendment of the U.S. *Constitution* does not protect speech that involves threats, targeted harassment, and imminent danger through incitement of violence, it does protect freedom of speech, no matter hOW offensive, distasteful or bigoted that speech might be. In fact, under U.S. law, there is no legal definition of unpatriotic speech [41]. Moreover, Section 230 of the 1996 U.S. *Communications Decency Act* offers significant protections to social media platforms, stating that “no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider” [42], meaning that platforms such as Facebook, Twitter, YouTube and Instagram cannot be held liable for user-generated content. In other words, it could prove difficult for the U.S. to criminalize the type of activity conducted by the Russian Internet Research Agency, without some major amendments to long-standing American legislation, and dramatic changes to legal precedent.

In Canada, the *Charter of Rights and Freedoms* states that individuals have the right to “freedom of thought, belief, opinion and expression, including freedom of the press and other media of communication” [43]. While actions such as defamatory libel and hate propaganda are prohibited by the *Canadian Criminal Code* [34], [44], the courts have gone to considerable lengths to protect freedom of expression. For example, in *Crouch v. Snell* [45] a case involving adult cyberbullying and the Province of Nova Scotia’s *Cyber-safety Act* [46], the judge confirmed that the right to freedom of expression “extends to any number of unpopular or distasteful expressions, including some forms of defamatory libel, hate propaganda and false news” [47]. *R. v. Elliott* [48], heard in 2016, was a criminal harassment case in which the accused repeatedly communicated with (and allegedly harassed) two feminist activists via Twitter, both at hashtags which they had created, and at hashtags with which they were affiliated. The judge opined that Twitter was like a “public square,” observing that creating a hashtag where people could follow you was similar to “announcing a public meeting,” further stressing that the fact that some opinions may be “morally offensive” to some people is *not* criminal.

Evidently, contentious issues involving freedom of expression and freedom of opinion can be expected to limit any effort to regulate the publication of “fake news” on social media. To be effective, regulatory agencies may need to target the creation of fraudulent Facebook pages and Twitter feeds, and in addition, the use of social bots that amplify messages in order to create the false impression that the messages have more followers and interactions than they do in reality.

V. CONCLUSION

With democracy under threat from the intentional (and perhaps criminal) manipulation of Cloud-based social media, and the resultant digital wildfires [49], legislators, regulators and service providers are eagerly seeking solutions and defenses against disinformation warfare. We have described the brazen attempts by the Russian Internet Research Agency to manipulate public opinion in the U.S. and U.K., wherein the use of so-called "fake news" sought to influence democratic processes across international boundaries. Looking ahead to technological responses, we anticipate developing tools that will permit agencies to filter and identify suspicious social network content. While subject to further research and verification, our reported 99.8 percent accuracy in classifying "fake news" demonstrates the feasibility of this objective. Yet, in turn, such developments may infringe upon the privacy and personal rights of the individual. Free speech and data privacy need to be balanced against the requirements for management and control of disinformation threats, but such balance is not easily achieved. Indeed, there is a fine line between the monitoring of social media and the potential abrogation of the right to privacy, to the extent that such privacy rights are believed to exist in the public domain. This conflict is evident from the legislative efforts that we have considered from the U.K, Europe, the U.S. and Canada. In each jurisdiction, there is a marked tension between these conflicting rights under the law. The clear conclusion is that responses from legislators and regulators to the type of weaponization of Cloud-based social media witnessed during the U.S. Presidential election and the Brexit referendum will impact widely upon the liberty of individuals, and give rise to much contentious litigation in the years to come.

ACKNOWLEDGMENT

This research project would not have been possible without funding from the Cyber Security Cooperation Program, operated by the National Cyber Security Directorate of Public Safety Canada. We would also like to thank our research assistants, Soobin Rim and Aynsley Pescitelli for their assistance with the data.

REFERENCES

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

[2] A. Badawy, E. Ferrara and Lerman, K., "Analyzing the Digital Traces of Political Manipulation: The 2016 Russian Interference Twitter Campaign, *arXiv* preprint arXiv:1802.04291, 2018.

[3] M. T. Bastos and D. Mercea, "The Brexit botnet and user-generated hyperpartisan news," *Social Science Computer Review*, 0894439317734157, 2017.

[4] H. Allcott and M. Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives*, 31(2), pp. 211-236, 2017.

[5] R. S. Mueller III, "Report on the Investigation into Russian Interference in the 2016 Presidential Election, pp. 1-448, 2019. URL: www.justsecurity.org/wp-content/uploads/2019/04/Muellller-Report-Redacted-Vol-II-Released-04.18.2019-Word-Searchable.-Reduced-Size.pdf [Last Accessed: 2019.04.22]

[6] M. Field and M. Wright, "Russian trolls sent thousands of pro-Leave messages on day of Brexit referendum, Twitter data reveals: Thousands of Twitter posts attempted to influence the referendum and US elections," *The Telegraph*, 2018. URL: www.telegraph.co.uk/technology/2018/10/17/russian-iranian-twitter-trolls-sent-10-million-tweets-fake-news/ [Last accessed: 2019.04.8]

[7] G. Evolvi, "Hate in a Tweet: Exploring Internet-Based Islamophobic Discourses," *Religions*, *9*(10), pp. 37-51, 2018.

[8] T. Romm, "Facebook's Zuckerberg just survived 10 hours of questioning by Congress," *Washington Post*. URL: www.washingtonpost.com/news/the-switch/wp/2018/04/11/zuckerberg-facebook-hearing-congress-house-testimony/?utm_term=.f06997434776, April 11, 2018. [Last accessed: 2019.04.8]

[9] Politico Staff, "Full text: Mark Zuckerberg's Wednesday testimony to Congress on Cambridge Analytica," April 9, 2018. URL: https://www.politico.com/story/2018/04/09/transcript-mark-zuckerberg-testimony-to-congress-on-cambridge-analytica-509978 [Last accessed: 2019.04.8]

[10] A. Satariano, "Facebook Faces Tough Questions in Britain That It Avoided in the U.S.," 2018. URL: www.nytimes.com/2018/04/26/business/facebook-british-parliament.html [Last accessed: 2019.04.8]

[11] J. P. Tasker, "'We are sorry': Facebook execs grilled by Canadian MPs over Cambridge Analytica scandal: For 2 years, Facebook knew personal info of thousands of Canadians may have been in hands of a third party," *CBC News*, April 2018. URL: www.cbc.ca/news/politics/facebook-execs-house-of-commons-sorry-1.4626206 [Last accessed: 2019.04.8]

[12] D. Ebner and C. Freeze, "AggregateIQ, Canadian data firm at centre of global controversy, was hired by clients big and small," *Globe and Mail*, April, 2018. URL: www.theglobeandmail.com/canada/article-aggregateiq-canadian-data-firm-at-centre-of-global-controvery-was [Last accessed: 2019.04.8]

[13] C. Shao, P. M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczerand and G. L. Ciampaglia, "Anatomy of an online misinformation network," PloS one, 13(4), e0196087, 2018.

[14] H. Berghel, "Lies, damn lies, and fake news," *Computer*, 50(2), pp. 80-85, 2017.

[15] J. E. Kirtley, "Getting to the Truth: Fake News, Libel Laws, and 'Enemies of the American People,'" 2018. URL: www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-ongoing-challenge-to-define-free-speech/getting-to-the-truth/ [Last accessed: 2019.04.8]

[16] N. W. Jankowski, "Researching fake news: A selective examination of empirical studies," Javnost-The Public, 25(1-2), pp. 248-255, 2018.

[17] E. C. Tandoc Jr, Z. W. Lim and R. Ling, "Defining 'fake news': A typology of scholarly definitions," *Digital Journalism*, 6(2), pp. 137-153, 2018.

[18] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer and M. Schudson, "The science of fake news," *Science*, *359*(6380), pp. 1094-1096, 2018.

[19] E. Shearer and K. E. Matsa, "News Use Across Social Media Platforms 2018: Most Americans continue to get news on social media, even though many have concerns about its accuracy," Pew Research Center, 2018. URL: www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/ [Last accessed: 2019.04.8]

[20] N. Newman, R. Fletcher, and A. Kalogeropoulos, Reuters Institute Digital News Report 2018. URL: http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf?x89475 [Last accessed: 2019.04.8]

[21] C. Shao, P. M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer and G. L. Ciampaglia, "Anatomy of an online misinformation network," *PloS one*, *13*(4), e0196087, 2018.

[22] Central Intelligence Agency, Federal Bureau of Investigation and National Security Agency, "Assessing Russian Activities and Intentions in Recent US Elections," 2017. URL: www.dni.gov/files/documents/ICA_2017_01.pdf [Last accessed: 2019.04.8]

[23] United States v. Internet Research Agency LLC, Case 1:18-cr-00032-DLF, The United States District Court for the District Of Columbia, February 26, 2018. URL: www.justice.gov/file/1035477/download [Last accessed: 2019.04.8]

[24] V. Narayanan, P. N. Howard, B. Kollanyi and M. Elswah, "Russian involvement and junk news during Brexit," (2017). URL: comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/12/Russia-and-B rexit-v27. pdf [Last accessed: 2019.04.8]

[25] The Economist, "Russian Twitter trolls meddled in the Brexit vote. Did they swing it?," 2017. URL: www.economist.com/britain/2017/11/23/russian-twitter-trolls-meddled-in-the-brexit-vote-did-they-swing-it [Last accessed: 2019.04.8]

[26] M. Fox, "Fake News: Lies spread faster on social media than truth does," *NBC News*, 2018. URL: www.nbcnews.com/health/health-news/fake-news-lies-spread-faster-social-media-truth-does-n854896 [Last accessed: 2019.04.8]

[27] D. L. Linvill and P. L. Warren, "Troll factories: The Internet Research Agency and state-sponsored agenda-building," Resource Centre on Media, 2018.

[28] G. Weir, R. Frank, B. Cartwright and E. Dos Santos, "Positing the problem: enhancing classification of extremist web content through textual analysis," International Conference on Cybercrime and Computer Forensics (IEEE Xplore), June 2016.

[29] G. Weir, K. Owoeye, A. Oberacker and H. Alshahrani, "Cloud-based textual analysis as a basis for document classification," *International Conference on High Performance Computing & Simulation (HPCS)*, pp. 672-676, July 2018.

[30] M. Hall, E. Frank, H. Geoffrey, B. Pfahringer, P. Reutemann and I. Witten, "The Weka data mining software: an update," *SIGKDD Explorations*, vol. 11, pp. 10-18, 2009.

[31] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, pp. 5-32, 2001.

[32] UN General Assembly. *Universal declaration of human rights*. 1948. URL: http://www.un.org/en/udhrbook/pdf/udhr_booklet_en_web.pdf [Last accessed: 2019.04.8]

[33] J. Walker, *Hate Speech and Freedom of Expression: Legal Boundaries in Canada*, Library of Parliament, 2018. URL: lop.parl.ca/sites/PublicWebsite/default/en_CA/ResearchPublications/201825E [Last accessed: 2019.04.8]

[34] Council of the European Union, *Council Framework Decision 2008/913/JHA of 28 November 2008 on Combating Certain Forms and Expressions of Racism and Xenophobia by Means of Criminal Law*, 2008. URL: publications.europa.eu/en/publication-detail/-/publication/f015ed06-b071-41e1-84f1-622ad4ec1d70 [Last accessed: 2019.04.8]

[35] Article 19, *United Kingdom (England and Wales): Responding to 'hate speech*,' 2018. URL: www.article19.org/wp-content/uploads/2018/06/UK-hate-speech_March-2018.pdf [Last accessed: 2019.04.8]

[36] J. Garlandand N. Chakraborti, "Divided by a common concept? Assessing the implications of different conceptualizations of hate crime in the European Union," *European Journal of Criminology*, *9*(1), pp. 38-51, 2012.

[37] Judiciary of Scotland, *PF v Mark Meecham*, 2018. URL: http://www.scotland-judiciary.org.uk/8/1962/PF-v-Mark-Meechan [Last accessed: 2019.04.8]

[38] K. Malik, "The 'Nazi pug': giving offence is inevitable and often necessary in a plural society," *The Guardian*, March 2018. URL: www.theguardian.com/commentisfree/2018/mar/25/being-offensive-should-not-be-illegal-in-society-that-defends-free-speech [Last accessed: 2019.04.8]

[39] E. McGuire, "Free Speech is Dead in Britain," *The American Spectator*, March 2018. URL: spectator.org/free-speech-is-dead-in-britain/ [Last accessed: 2019.04.8]

[40] L. J. Gross and J. David, *Munim Abdul and Others v Director of Public Prosecutions*, EWHC 247 (Admin), 2011. URL: swarb.co.uk/abdul-and-others-v-director-of-public-prosecutions-admn-16-feb-2011/ [Last accessed: 2019.04.8]

[41] American Library Association, "Hate speech and hate crime" nd. URL: http://www.ala.org/advocacy/intfreedom/hate [Last accessed: 2019.04.8]

[42] *Communications Decency Act*, 47 U.S.C. §230, 1996. URL: http://www.columbia.edu/~mr2651/ecommerce3/2nd/statutes/CommunicationsDecencyAct.pdf [Last accessed: 2019.04.8]

[43] *Canadian Charter of Rights and Freedoms*, s8, Part 1 of the *Constitution Act*, 1982, being Schedule B to the Canada Act 1982 (UK), c 11, 1982.

[44] *Criminal Code*, RSC 1985, c C-46, s 318(1)(a), 1985.

[45] G. G. McDougall, *Crouch v. Snell*, vol. 2015 NSSC 340, 2015.

[46] Nova Scotia Government, *Cyber-Safety Act: An Act to Address and Prevent Cyberbullying*, vol. 61, 2013.

[47] B. Cartwright,"Cyberbullying and 'The Law of the Horse': A Canadian viewpoint," *Journal of Internet Law*, *20*(10), pp. 14–26, 2017.

[48] B. Knazan, *R. v. Elliott*, vol. [2016] ONCJ 310, 2016.

[49] H. Webb, P. Burnap, R. Procter, O. Rana, B. C. Stahl, M. Williams, … M. Jirotka., "Digital Wildfires: Propagation, Verification, Regulation, and Responsible Innovation," *ACM Transactions on Information Systems*, *34*(3), pp. 15:1–15:23, 2016.