Rise of the Troll: Exploring the Constitutional Challenges to Social Media and Fake News Regulation in the Philippines

Jomari James T. De Leon[*]
Keir Cedric L. Enriquez[**]
Jose Angelo C. Tiglao[***]

I. INTRODUCTION 151
A. Background of the Study
B. Significance of the Study
II. SOCIAL MEDIA AS A CONCEPT 162
A. The Fake News Phenomenon
B. Understanding the Concept of Echo Chambers
C. Rise of the Trolls
III. SOCIAL MEDIA REGULATION 172
A. Challenges of Social Media Regulation
B. Constitutionality of Social Media Regulation
IV. EXISTING REGULATORY FRAMEWORKS 187
A. The International Regulatory Framework for Social Media
B. The Philippines' Regulatory Framework for Social Media
C. Voluntary Self-Regulating Framework of Social Media Companies
D. Limitations of Self-Imposed Measures

---

* '18 J.D., De La Salle University College of Law. He ranked 7th in his batch when he graduated from De La Salle University College of Law.

** '18 J.D., De La Salle University College of Law.

*** '18 J.D., De La Salle University College of Law. The Author is currently a Junior Associate in Quisumbing Torres, a member firm of Baker McKenzie. He ranked 6th in his batch when he graduated from De La Salle University College of Law.

The Authors are the recipients of the Most Outstanding Thesis Award in 2018 as conferred by the De La Salle University College of Law. This Article is an abridged version of the Authors' Juris Doctor thesis.

Cite as 64 ATENEO L.J. 150 (2019).

V. LAW AND JURISPRUDENCE ON FREEDOM OF EXPRESSION ... 204
A. The Danger of Fake News in Philippine Society
VI. THE LIABILITY OF ONLINE INTERMEDIARIES 215
VII. COMPARATIVE ANALYSIS: THE VARIOUS STATE APPROACHES IN ADDRESSING THE FAKE NEWS PROBLEM 222
VIII. DETERMINING THE BEST APPROACH IN ADDRESSING THE FAKE NEWS PHENOMENON IN THE PHILIPPINE SOCIETY 227
A. Punishing the Architects of Disinformation
B. Online Intermediary Liability and Accountability
C. Closer Look on the Alternative: "Notice and Correct" Procedure
IX. CONCLUSION AND RECOMMENDATIONS 238

I. INTRODUCTION

A. Background of the Study

Social media has played an active role in shaping the Philippine society over the past few years.[1] Social networking sites such as Facebook, Twitter, and YouTube have been constantly used to shape public opinion, initiate movements, and champion causes in various parts of the world.[2] It can be said that the list of advantages introduced by social media is unlimited.[3] It has connected the world in a very fast pace because a person can do almost anything with the click of a button. It is its capacity "to share information,

1. *See* Alexandra Guzman, 6 ways social media is changing the world, *available at* https://www.weforum.org/agenda/2016/04/6-ways-social-media-is-changing-the-world (last accessed July 25, 2019).

2. *See* Monica Anderson, et al., 1. Public attitudes toward political engagement on social media, *available at* https://www.pewinternet.org/2018/07/11/public-attitudes-toward-political-engagement-on-social-media (last accessed July 25, 2019).

3. *See* Guzman, *supra* note 1.

ideas, personal messages, and other content"[4] across the globe that makes it very powerful.

Likewise, the power of social media and its effects are far-reaching — useful to the informed and dangerous to the irresponsible. Not much can be said to those who wield this platform in the responsible exercise of their right to free speech. Social media was monumental in the success of various global movements, such as the Arab Spring.[5] There, social networking sites were used as a platform to cascade messages about freedom and democracy to help raise expectations for a successful political uprising.[6] For example, prior to the resignation of Egyptian president Hosni Mubarak, the total rate of tweets about political change "ballooned from 2,300 a day to 230,000 a day."[7]

Another example was the role of social media and its users during the London Riots in 2011.[8] In that instance, the riots were coordinated through Blackberry Messenger and Twitter, among others.[9] It resulted to concerted efforts by the people to show their frustration to the police due to the unfortunate killing of Mark Duggan.[10]

Finally, who can forget what happened during Occupy Wall Street, where more than 450,000 Facebook users went to the streets during the movement?[11]

---

4. Merriam-Webster, Inc., Social Media, *available at* https://www.merriam-webster.com/dictionary/social%20media (last accessed July 25, 2019).

5. *See* Catherine O'Donnell, New study quantifies use of social media in Arab Spring, *available at* http://www.washington.edu/news/2011/09/12/new-study-quantifies-use-of-social-media-in-arab-spring (last accessed July 25, 2019).

6. O'Donnell, *supra* note 5.

7. *Id.*

8. *See* Christian Fuchs, *BEHIND THE NEWS: Social media, riots, and revolution*, 36 CAPITAL & CLASS 383, 383-85 (2012).

9. Fuchs, *supra* note 8, at 384-85.

10. *Id.* at 383.

11. Craig Kanalley, Occupy Wall Street: Social Media's Role In Social Change, *available at* https://www.huffpost.com/entry/occupy-wall-street-social-media_n_999178 (last accessed July 25, 2019).

Undoubtedly, these platforms have been used and will continue to be used as avenues of change and ideas.

It is because of these advantages that there are people who have chosen to abuse it for their personal agenda. The sheer number of social media users alone — around 2.51 billion users in 2017[12] — is a testament to how much change or damage it can cause upon a person, institution, or idea, if used improperly.

Now comes the question: is there a need for protection from the adverse effects of social media? The right to free speech is considered as one of the most important civil liberties, to which any form of regulation is frowned upon by the law.[13] Throughout this Article, the right to free speech in social media shall be placed closely beside the need of protection from its abusive nature.

Several countries have taken the lead in regulating social media freedom. Germany, for example, has recently passed a law imposing stricter standards for those who post, spread, and create news on these platforms.[14] This measure was supported by Josef Schuster, the president of the Central Council of Jews in Germany, who said, "We do not want an internet police or thought control[.]"[15] He then added, "But when hatred is stoked, and the legal norms in our democracy threaten to lose their relevance, then we need to intervene."[16]

---

12. Nearly One-Third of the World Will Use Social Networks Regularly This Year, *available at* https://www.emarketer.com/Article/Nearly-One-Third-of-World-Will-Use-Social-Networks-Regularly-This-Year/1014157 (last accessed July 25, 2019)

13. The Diocese of Bacolod v. Commission on Elections, 747 SCRA 1, 83-84 (2015) (citing Reyes v. Bagatsing, 125 SCRA 553, 563 & 570 (1983) & Blo Umpar Adiong v. Commission on Elections, 207 SCRA 712, 712, 715, & 717 (1992)).

14. *See* Eric Auchard & Hans-Edzard Busemann, Germany plans to fine social media sites over hate speech, *available at* http://www.reuters.com/article/us-germany-fakenews-idUSKBN16L14G (last accessed July 25, 2019).

15. *Id.*

16. *Id.*

The same can be said with the results of the 2016 United States (U.S.) Presidential Elections.[17] According to various reports, social media was employed to proliferate fake news in order to shift and influence public opinion with respect to the candidates.[18] Similarly, the Philippines has experienced much of this proliferation even beyond the national presidential elections.[19] It is because of this that several lawmakers, such as Rep. Pantaleon Alvarez and Sen. Grace Poe, were motivated to file their respective bills in both Houses of Congress seeking to either regulate the use of social media in the Philippines[20] or prohibit public officers from publishing or disseminating fake news.[21]

17. *See generally* Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, J. ECON. PERSPECTIVES, Volume No. 31, Issue No. 2, at 211.

18. Allcott & Gentzkow, *supra* note 17, at 212 (citing Craig Silverman, This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook, *available at* https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook (last accessed July 25, 2019) & Craig Silverman & Jeremy Singer-Vine, Most Americans Who See Fake News Believe It, New Survey Says, *available at* https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey (last accessed July 25, 2019)).

19. See Maria Pilar M. Lorenzo, The rise of fake news: the Philippine case, *available at* https://policyblog.uni-graz.at/2017/11/the-rise-of-fake-news-the-philippine-case (last accessed July 25, 2019).

20. Rose-An Jessica Dioquino, House bill to criminalize fake Facebook, other social media accounts, *available at* http://www.gmanetwork.com/news/story/600476/news/nation/house-bill-to-criminalize-fake-facebook-other-social-media-accounts (last accessed July 25, 2019). *See generally* An Act Regulating the Use of Social Media, Prescribing Penalties and for Other Purposes, H.B. No. 5021, 17th Cong., 1st Reg. Sess. (2017).

21. Paolo Romero, *Palace, government offices responsible in fighting fake news — Poe*, PHIL. STAR, Jan. 31, 2018, *available at* https://www.philstar.com/headlines/2018/01/31/1783136/palace-government-offices-responsible-fighting-fake-news-poe (last accessed July 25, 2019). *See generally* An Act Amending Sections 4 (b) and 7 of Republic Act No. 6713, Otherwise Known as the "Code of Code of Conduct

This move by Philippine lawmakers is not shocking considering the its political climate. In a recent study entitled *Digital in 2017*, the Philippines was ranked first in terms of the number of hours spent by a person in social media.[22] There is evidence that since Filipinos spend so much time in social networking sites, there is a high exposure to news, whether real or fake. Further, it is not just a question of genuine news stories, but also the authenticity of those behind these accounts because many have alleged that the same are bogus.[23] Hence, the question: are all of these accounts real? Is it possible that these accounts — which spread false or fake news — are fake?

Maria A. Ressa, Chief Executive Officer of Rappler, thinks so.[24] In her article entitled *Propaganda war: Weaponizing the Internet*, she points out that her company discovered that many of those part of Duterte's online campaign machinery were fake accounts, bots, and trolls.[25] She calls this out as the creation of a "manufactured reality."[26] She says, "What [we are] seeing on social media again is manufactured reality[.] They also create a very real

---

and Ethical Standards for Public Officials and Employees", and for Other Purposes, S.B. No. 1680, 17th Cong., 2d Reg. Sess. (2018).

22. Miguel R. Camus, *PH world's No. 1 in terms of time spent in social media*, PHIL. DAILY INQ., Jan. 24, 2017, *available at* http://technology.inquirer.net/58090/ph-worlds-no-1-terms-time-spent-social-media (last accessed July 25, 2019).

23. *See* Butch Fernandez, *Facebook officials uncover, removes 583,000 fake accounts, SP Sotto reports*, BUS. MIRROR, Oct. 9, 2018, *available at* https://businessmirror.com.ph/2018/10/09/facebook-officials-uncover-removes-583000-fake-accounts-sp-sotto-reports (last accessed July 25, 2019) & Consuelo Marquez, *FB cracks down on fake accounts, fake news to keep PH 'election integrity'*, PHIL. DAILY INQ., Jan. 24, 2019, *available at* https://technology.inquirer.net/82999/fb-cracks-down-on-fake-accounts-fake-news-to-keep-ph-election-integrity#ixzz5s1MDyFJe (last accessed July 25, 2019).

24. Maria A. Ressa, Propaganda war: Weaponizing the internet, *available at* http://www.rappler.com/nation/148007-propaganda-war-weaponizing-internet (last accessed July 25, 2019).

25. *Id.*

26. BBC Trending, Trolls and triumph: A digital battle in the Philippines, *available at* http://www.bbc.com/news/blogs-trending-38173842 (last accessed July 25, 2019).

chilling effect against normal people, against journalists [who] are the first targets[.]"[27]

Based on her statement, the misuse of social media in the Philippines has the potential to create a chilling effect in society [28] No one is spared from such blunder — not even the Vice President of the Philippines, Maria Leonor "Leni" Robredo.[29] Recently, she has been the victim of various social media attacks, to which her daughters have also been subject to malicious remarks.[30] During one of her speeches in Cebu, she declared a "war on online trolls" due to the intense proliferation of "lies, fake news, or alternative facts[,]"[31] which she said if not stopped, will "assume the appearance of truth."[32] The power of social media is being used to advance political agenda in the Philippines, in the guise of free exercise of speech. Arguably, this is not the intent of the Constitution.

The lack of regulation of social media use has resulted in a unique yet dangerous online environment. Gone are the days when cybercrime prevention laws were sufficient to protect private and public interests. Nowadays, these platforms are being used not only to attack an individual directly, but also to spread lies, fake news, and alternative facts, which aim to sway public opinion so that they, in turn, shall go against the individual, institution, or idea.

Therefore, the need to regulate those who use social media as a veil to escape liability becomes apparent. Specifically, those who hide behind their technology and create fake accounts, trolls, and bots in order to spread lies.

---

27. *Id.*

28. *Id.*

29. *See* Nikko Dizon, *Robredo tells of online rape threats on daughters*, PHIL. DAILY INQ., Mar. 10, 2017, *available at* http://newsinfo.inquirer.net/879389/robredo-tells-of-online-rape-threats-on-daughters (last accessed July 25, 2019).

30. Dizon, *supra* note 29.

31. DJ Yap & Izobelle T. Pulgo, *Robredo declares war on online trolls*, PHIL. DAILY INQ., Mar. 4, 2017, *available at* https://newsinfo.inquirer.net/877374/robredo-declares-war-on-online-trolls (last accessed July 25, 2019).

32. *Id.*

However, the idea of regulating social media may seem impossible to many, especially to the international community.

Any form of domestic regulation to free speech must be weighed against the State's international obligations under the Universal Declaration of Human Rights (UDHR) — the generally agreed foundation of international human rights law.[33] Clearly, the rights under this declaration bind the Filipino people because they are recognized under the Constitution.[34]

The UDHR "makes it clear that all human rights are indivisible and interrelated, and that equal importance should be attached to each and every right."[35] Article 19 provides, "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive[,] and impart information and ideas through any media and regardless of frontiers."[36]

Regulation, through means like identity verification of possible users, raises complicated questions and issues as to what should and should not be regulated. Obviously, it raises questions regarding the right to free speech protected under the UDHR.

Thus,

> [t]he U.N. recognizes that all individuals should express their thoughts and ideas in an open and unrestricted environment. As a result, when anyone posts something on social media, there is an expectation that this is protected speech. When speech is protected, there is a high standard that the government must meet in order to justify censorship. ... [S]creening and

33. *See* Universal Declaration of Human Rights, G.A. Res. 217 A (III), art. 19, U.N. Doc. A/RES/3/217 A (Dec. 10, 1948) [hereinafter UDHR].

34. PHIL. CONST. art. 2, § 2.

35. Office of the United Nations High Commissioner for Human Rights, The United Nation Human Rights Treaty System: An introduction to the core human rights treaties and the treaty bodies (Fact Sheet No. 30 of the Human Rights Fact Sheet Series Published by the U.N.) at 2, *available at* http://www.ohchr.org/Documents/Publications/FactSheet30en.pdf (last accessed July 25, 2019).

36. UDHR, *supra* note 33, art. 19.

> removal of offensive social media [posts or accounts] raises complicated issues over what should and should not be censored.[37]

Yet, the U.N. recognized that the right is not without limitation.[38] In the second paragraph of Article 29 of the UDHR, it states, "In the exercise of his rights and freedoms, everyone shall be subject only to such limitations as are determined by law solely for the purpose of securing due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order[,] and the general welfare in a democratic society."[39]

Aside from the UDHR, for regulation to pass, it must adhere to the International Covenant on Civil and Political Rights (ICCPR).[40]

> [T]he [UDHR] was the first attempt by all States to agree, in a single document, on a comprehensive catalogue of the rights of the human person. As its name suggests, it was not conceived of as a treaty but rather a proclamation of basic rights and fundamental freedoms, bearing the moral force of universal agreement.[41]

Thus, in 1966, the U.N. General Assembly adopted the ICCPR, which would be directly binding upon States that agreed and ratified its terms.[42]

Any form of regulation, especially one that involves the right to free speech, may run contrary to Article 19 of the ICCPR, which states, "Everyone

---

37. Paulina Wu, *Impossible to Regulate: Social Media, Terrorists, and the Role for the U.N.*, 16 CHI. J. INT'L L. 281, 290 (2015).

38. *Id.*

39. *Id.* (citing UDHR, *supra* note 33, art. 29, ¶ 2).

40. *See* International Covenant on Civil and Political Rights art. 19, *opened for signature* Dec. 19, 1966, 999 U.N.T.S. 171 [hereinafter ICCPR].

41. Office of the United Nations High Commissioner for Human Rights, Civil and Political Rights: The Human Rights Committee (Fact Sheet No. 15 of the Human Rights Fact Sheet Series Published by the U.N.) at 1, *available at* http://www.ohchr.org/Documents/Publications/FactSheet15rev.1en.pdf (last accessed July 25, 2019).

42. *Id.*

shall have the right to hold opinions without interference"[43] and "shall have the right freedom of expression."[44]

However, the same Article states that such right may be subject to certain restrictions, "but these shall only be such as are provided by law and are necessary: (a) [f]or respect of the rights or reputations of others; [and] (b) [f]or the protection of national security or of public order ... , or of public health or morals."[45]

This very high standard of constitutional protection to a person's right to free speech is obviously present in the Philippines.[46] Nonetheless, for the past few years, recent developments to the country's body of laws have been introduced to regulate and punish those who abuse their right. In the dawn of social media, the existing penal provisions on libel under the Revised Penal Code have found applicability in the internet sphere through the passage of the Cybercrime Prevention Act.[47] Thus, it is clear that in the Philippine jurisdiction, regulation is possible.

A defamatory allegation is one that "ascribes to a person [possession of a vice or defect,] commission of a crime, real or imaginary[,] or any act, omission, condition, status, or circumstance [having the tendency] to dishonor[,] discredit[,] or put [the person] in contempt, or [having the tendency] to blacken the memory of [a] dead [one]."[48] To determine a defamatory statement, all the words that are used must be construed in their entirety or as a whole.[49] They should be taken in their ordinary meaning as would "be understood by persons reading them, [except if] it [seems] that they were [understood] and [used] in another sense."[50] Freedom of expression

43. ICCPR, *supra* note 40, art. 19, ¶ 1.

44. *Id.* art. 19, ¶ 2.

45. *Id.* art. 19, ¶ 3.

46. *See* PHIL. CONST. art. III, § 4.

47. *See* An Act Defining Cybercrime, Providing for the Prevention, Investigation, Suppression and the Imposition of Penalties Therefor and for Other Purposes [Cybercrime Prevention Act of 2012], Republic No. 10175, § 4 (b) (4) (2012).

48. Lopez v. People, 642 SCRA 668, 679 (2011).

49. *Id.* at 679-80 (citing Buatis, Jr. v. People, 485 SCRA 275, 286 (2006)).

50. *Lopez*, 642 SCRA at 679-80 (citing *Buatis, Jr.*, 485 SCRA at 286).

cannot be used to broadcast lies or even half-truth.[51] This is not consistent with "[observing] honesty and good faith."[52] It is not a tool to "insult others, [to] destroy ... name or reputation[,] or [to] bring [a person] into disrepute."[53] This is contrary to acting with justice and giving everyone his or her due.[54]

While libel has been recognized under the Revised Penal Code for a long time,[55] the provisions of the Cybercrime law have just recently passed constitutional scrutiny. The Supreme Court in *Disini, Jr. v. Secretary of Justice*,[56] ruled that online activities can be subject to State regulation.[57] Defamatory statements made in cyberspace can now be prosecuted as cyber libel.[58] However, the Philippines has no law that protects those who suffer damage from the proliferation of lies, fake news, and alternative facts in the guise of false accounts and trolls. The only existing measure to combat such is provided by the networking sites themselves, such as the report option on Twitter.[59]

Indeed, the author of the alleged defamatory online statement may be prosecuted if he or she is a real person. Unfortunately, current events show that those who spread these defamatory statements are usually accounts using fake identities.[60] This has led to excessive bashing, flaming, and hate — all of which may arguably be subject of regulation.

---

51. In Re: Emil P. Jurado, 243 SCRA 299, 325 (1995).

52. *Id.*

53. *Id.*

54. *Id.*

55. *See* Disini, Jr. v. Secretary of Justice, 723 SCRA 109, 131 (2014) (citing Worcester v. Ocampo, 22 Phil. 42 (1912)).

56. Disini, Jr. v. Secretary of Justice, 716 SCRA 237, 320 (2014).

57. *Id.* at 320 & 354-56.

58. *Id.* at 320.

59. *See, e.g.*, Report Violations, *available* at https://support.twitter.com/articles/15789 (last accessed July 25, 2019).

60. *See* Disini & Disini Law Office, Internet Trolls in the Philippines, *available at* https://elegal.ph/internet-trolls-in-the-philippines (last accessed July 25, 2019).

Nowadays, by just looking at any social networking site, the proliferation of these fake accounts is apparent — more so, the spread of false information. In one account, a certain Mr. Punzalan was wrongfully linked as the suspect in a controversial road rage killing through an article published by Top Gear Philippines.[61] Punzalan, who was later found to be not a suspect, suffered great fear and trauma due to the online article for fear for his life.[62] Clearly, Top Gear may be liable under the crime of cyber libel. However, one cannot deny that equally as damaging were the comments made against him, which can be real or not. If real, then it is easy — one can simply file a criminal action against the person. If fake, what is the remedy? Will the State not intervene and prevent further damage due to the proliferation of these fake accounts? Clearly, this is an issue that needs to be resolved in the dawn of technological advancement.

The Article aims to determine if the publication and/or dissemination of fake news on social media is a speech protected by the Constitution, and if not, what kind of regulation is best applicable.

As regards these social media companies, this Article seeks to identify how it should be treated in light of the culture of disinformation and how users have been exploiting the features of the company to spread fake news.

B. Significance of the Study

Social media empowers users to exercise their freedom of expression. It encourages the free discussion of ideas through the Internet sans any form of restriction or regulation. Because of this, social media has become very powerful to the point that it poses a certain level of danger to society, provided it is abused by bad actors.

Its integration to the average person's life is unquestionable based on *Digital in 2017 Global Overview*, where Filipinos, in particular, are said to spend most of their time in social media — compared to any other nationality in the

61. Niko Baua, Cyberbullying victim sues Top Gear PH editorial board, *available at* http://news.abs-cbn.com/news/08/12/16/cyberbullying-victim-sues-top-gear-ph-editorial-board (last accessed July 25, 2019).

62. *Id.*

world — with an average use of 4.17 hours daily in 2016.[63] Globally, there are 2.8 billion active social media users[64] — all of which can use it for their personal, commercial, or even, political use. This growth is outstanding and deserving of praise as a symbol of technological advancement and innovation in the 21st century.

Nonetheless, this Article recognizes that amidst all the positive things, the problems brought by the rise of fake news, echo chambers, and social media bots are too important not to give attention to. There may have been a time when these problems were so miniscule that it could have been easily ignored, but the changing landscape has proven that these three problems have caused drastic effects — from the results of a domestic election to the problems caused by it to an ordinary person. If not addressed immediately, these problems may threaten the country's democracy and the way Filipinos live their lives.

The Article seeks to explore the possible implications of social media regulation with respect to one's constitutional rights, specifically the right to free speech. It is very important to determine how the State should treat speech made on these social media companies and how it must regulate the actions of the platform itself. The danger lies in striking the balance because if the State is to be restrictive, it will open itself to censorship, while if the treatment is relaxed, it is tantamount to allowing the culture of impunity spread through the proliferation of fake news, which will most likely result to greater harms to society.

II. SOCIAL MEDIA AS A CONCEPT

We are in a generation where a significant number of people have access to at least one form of social media. The advancement of current technology has led to a more globalized world because through the mere click of a button, the thoughts of a person in one place can be easily shared across the globe. For the purposes of this Article, the Authors borrow the generic definition of social media, to wit — "forms of electronic communication ... through which users create online communities to share information, ideas, personal messages, and

63. Simon Kemp, Digital in 2017: Global Overview, *available at* https://wearesocial.com/sg/blog/2017/01/digital-in-2017-global-overview (last accessed July 25, 2019).

64. *Id.*

other content."[65] Today, the most well-known social networking sites are Facebook[66] and Twitter.[67]

Similar to existing media platforms, social media allows its users to spread content via the Internet. Due to its fast paced and accessible nature, it becomes readily available to almost everyone, as long as he or she has access to the Internet. It is because of this that social media is "often used for breaking news or sharing information of immediate importance."[68]

History provides several examples of how social media has shaped the way content is shared and used. On the one hand, in 2011, the civil war in Libya and the various occurrences within the state were disseminated to the global community through the use of social media.[69] On the other hand, Syrian refugees have used social media to seek for aid through spreading awareness by posting online the things happening in their country.[70] It seems that people opt to use social media to express their thoughts and sentiments because it proves to be very effective since it "requires little effort on the part of [the] followers or activists to engage with others and share information."[71]

65. Merriam-Webster, Inc., *supra* note 4.

66. Facebook, Company Info, *available at* https://newsroom.fb.com/company-info (last accessed July 25, 2019).

67. Twitter, Inc., Company, *available at* https://about.twitter.com/company (last accessed July 25, 2019).

68. Anne Herzberg & Gerald M Steinberg, *IHL 2.0: Is There a Role for Social Media in Monitoring and Enforcement?*, 45 ISR. L. REV. 493, 496 (2012).

69. Neal Ungerleider, Libya, YouTube, and the Internet, *available at* http://www.fastcompany.com/1731395/libya-youtube-and-internet (last accessed July 25, 2019).

70. Aryn Baker, For Syrians, Social Media is More Useful than the U.N. Security Council, *available at* http://time.com/35826/syria-u-n-soeial-media-yarmouk (last accessed July 25, 2019).

71. Herzberg & Steinberg, *supra* note 68, at 496.

Social media's content, unlike other media technologies such as television[72] or newspapers, is not subject to third-party regulation such as fact checking and filtering. This leads to users being able to post anything online without any form of screening prior to posting. Subsequently, this leads to information that may or may not be reliable.

An example of this feature is in the proliferation of fake news and terrorism. In a 2016 study about the U.S. presidential election, it was found that fake news favoring Donald Trump was shared a total of 30 million times on Facebook, while those that favored Hillary Clinton were shared under 8 million times.[73]

Insofar as terrorism is concerned, it is said that terrorists have "moved their online presence to YouTube, Twitter, Facebook and Instagram, and other social media outlets."[74] This seems to be an ongoing trend considering that "about 90[%] of organized terrorism on the Internet is being carried out through social media."[75] Social media's accessibility has led to the possibility that any message, including those belonging from terrorists, can be easily communicated to others.

The past few years have significantly paved the way for the development of social media. Nonetheless, these changes have given rise to several incidents concerning social media users who have abused their rights online by sharing

72. *See generally* ARTICLE 19, LONDON AND CMFR, MANILA, FREEDOM OF EXPRESSION AND MEDIA IN THE PHILIPPINES 44-45 (2005).

73. Allcott & Gentzkow, *supra* note 17, at 212.

74. Gabriel Weimann, New Terrorism and New Media (An Article Published Online by the Commons Lab of the Woodrow Wilson International Center for Scholars As Part of a Research Series) at 1, *available at* https://www.wilsoncenter.org/sites/default/files/STIP_140501_new_terrorism_F_0.pdf (last accessed July 25, 2019).

75. CBC News, Terrorist groups recruiting through social media, *available at* https://www.cbc.ca/news/technology/terrorist-groups-recruiting-through-social-media-1.1131053 (last accessed July 25, 2019).

harmful content or by engaging in trolling, online harassment, and the dissemination of fake news.[76]

Based on a recent survey, it was found that about four out of 10 Americans have been harassed on social media.[77] Online harassment can be traced from various sources, one of which is trolling.[78] In the realm of social media, the concept of trolling is not uncommon. A troll can be any person at any given time because, generally, people do not start off as trolls.[79] Over time, more people discovered the effect of trolling and the ways it shapes various aspects of life. Hence, there is a growing number of cases where trolls are made for that purpose alone — to troll.[80]

Trolling can cause damage to others. In 2012, Anita Sarkeesian started a Kickstarter campaign to seek funds for an online campaign involving

76. *See* Lee Rainie, et al., The Future of Free Speech, Trolls, Anonymity and Fake News Online, *available at* https://www.pewinternet.org/2017/03/29/the-future-of-free-speech-trolls-anonymity-and-fake-news-online (last accessed July 25, 2019) & Maeve Duggan, Online Harassment 2017, *available at* https://www.pewinternet.org/2017/07/11/online-harassment-2017 (last accessed July 25, 2019).

77. Duggan, *supra* note 76.

78. Maeve Duggan, Online Harassment: Introduction, *available at* https://www.pewinternet.org/2014/10/22/introduction-17 (last accessed July 25, 2019).

79. *See* Gaia Vince, Why good people turn bad online – and how to defeat your inner troll, *available at* https://www.independent.co.uk/news/long_reads/online-trolls-digital-societies-science-mary-beard-twitter-facebook-hate-speech-a8279596.html (last accessed July 25, 2019).

80. Justin Cheng, et al., Why people troll, according to science, *available at* http://www.businessinsider.com/find-out-why-any-of-us-are-capable-of-trolling-2017-3 (last accessed July 25, 2019).

misogyny.[81] During this time, she started receiving bomb and rape threats from people trolling her online.[82]

Social media bots, or those accounts created primarily to do something online such as trolling, have surfaced all across the internet.[83] A recent study showed that the bots' presence had a negative impact on the 2016 U.S. presidential elections because their acts "can potentially alter public opinion and endanger the integrity of the [presidential] election."[84]

The Philippines is not very far behind insofar as the presence of trolls and online bots are concerned. During the recently concluded presidential election, news reports stated that some "Duterte campaign insiders admitted that they used trolls or fake accounts"[85] during the campaign period. In short, people were paid to make fake accounts and troll on various social media accounts in order to shape public opinion.[86]

It is apparent that the use of social media has given rise to new ways of exercising one's right to free speech. However, it also gave birth to a new avenue for a person to spread propaganda, fake news, false information, and the like — all of which are threats to democracy.

---

81. Joel Stein, How Trolls Are Ruining the Internet, *available at* http://time.com/4457110/internet-trolls (last accessed July 25, 2019).

82. *Id.*

83. *See* Stein, *supra* note 81 & Donara Barojan, Understanding bots, botnets and trolls, *available at* https://ijnet.org/en/story/understanding-bots-botnets-and-trolls (last accessed July 25, 2019).

84. Alessandro Bessi & Emilio Ferrara, *Social bots distort the 2016 U.S. Presidential election online discussion*, FIRST MONDAY, Volume No. 21, Issue No. 11 (Authors' Abstract) (2016).

85. Chay F. Hofileña, Fake accounts, manufactured reality on social media, *available at* http://www.rappler.com/newsbreak/investigative/148347-fake-accounts-manufactured-reality-social-media (last accessed July 25, 2019).

86. *Id.*

A. The Fake News Phenomenon

The publication or dissemination of fake news on social media is an intentional and deliberate act committed by either individuals, groups, or organizations, who seek to propagate information that is either completely or partially false in order to shape public opinion or create controversy. More often than not, these fake news have a *grain of truth* in it, but this *truth* is highly twisted, exaggerated, without or taken out of context, and more.[87] If you look at Facebook, you can see that these fake news are usually published by websites which try to imitate genuine news publishers.[88] They try to pretend to be trustworthy in order to get their fake news across to users.[89] It is important to note that creators of fake news do not always seek to change public opinion; in fact, the goal of some is to divide society.[90]

Fake news' reach greatly increases the more times it is shared on the social media platform due to the algorithms of these social media companies. It is exactly because of these algorithms that the political climate has been greatly affected.[91] A highlighted example of fake news is the story apparently

87. Joshua Gillin, Fact-Checking fake news reveals how hard it is to kill pervasive 'nasty weed' online, *available at* http://www.politifact.com/punditfact/article/2017/jan/27/fact-checking-fake-news-reveals-how-hard-it-kill-p (last accessed July 25, 2019).

88. Richard Waters, et al., *Harsh truths about fake news for Facebook, Google and Twitter*, FIN. TIMES, Nov. 2, 2016, *available at* https://www.ft.com/content/2910a7a0-afd7-11e6-a37c-f4a01f1b0fa1 (last accessed July 25, 2019). *See, e.g.*, Christopher Elliott, Here Are The Real Fake News Sites, *available at* https://www.forbes.com/sites/christopherelliott/2019/02/21/these-are-the-real-fake-news-sites/#5ee0343c3c3e (last accessed July 25, 2019).

89. Emma Grey Ellis, Fake Think-Tanks Fuel Fake News—And the President's Tweets, *available at* https://www.wired.com/2017/01/fake-think-tanks-fuel-fake-news-presidents-tweets (last accessed July 25, 2019).

90. *See* SIMON HEGELICH, INVASION OF THE SOCIAL BOTS 3 (2016) & Charissa Yong, *Select Committee on fake news: Russian trolls divided societies and turned countries against one another*, STRAITS TIMES, Sep. 20, 2018, *available at* https://www.straitstimes.com/politics/select-committee-on-fake-news-russian-trolls-divided-societies-and-turned-countries-against (last accessed July 25, 2019).

91. Waters, et al., *supra* note 88.

published by the *Denver Guardian*, a newspaper which does not exist, that had claimed that the Federal Bureau of Investigation (FBI) agent who allegedly leaked the emails of Hillary Clinton during the U.S. elections had been murdered in a way that it was made to look like a suicide.[92] It was reported that this single article — fake news — was shared on Facebook numerous times, which led more users to view it and possibly, believe that it was true.[93]

The concept of fake news is not at all foreign. It is defined as those news which is "intentionally and verifiably false[ ] and could mislead readers."[94] A well-known example of fake news is the *Great Moon Hoax*, which was written and published by the *New York Sun* in 1835 detailing the alleged discovery of life on the moon.[95] A more recent example is the 2006 *Flemish Secession Hoax*, in which a Belgian public television station reported that the Flemish parliament had declared independence from Belgium — a report that a large number of viewers misunderstood as true.[96]

It is social media's dynamic and accessible nature that has caused the proliferation of fake news. Looking back, when the internet was not yet available, it was difficult to publish fake news because media entry barriers were high and to spread information normally entailed high costs.[97] But now, anyone can spread fake news as long as he or she has a social media account and has access to the internet.

---

92. *Id.*

93. *Id.*

94. Allcott & Gentzkow, *supra* note 17, at 213.

95. Craig Hlavaty, *'The Great Moon Hoax' of 1835 was one of the first examples of 'fake news'*, HOUSTON CHRONICLE, Dec. 12, 2016, *available at* http://www.chron.com/news/strange-weird/article/The-Great-Moon-Hoax-was-unleashed-onto-the-9184644.php (last accessed July 25, 2019).

96. Elliot Feldman, Flemish Secession Hoax, *available at* http://hoaxes.org/archive/permalink/flemish_secession_hoax (last accessed July 25, 2019).

97. James Carson, *Fake news: What exactly is it – and how can you spot it?*, TELEGRAPH, May 31 2019, *available at* http://www.telegraph.co.uk/technology/0/fake-news-origins-grew-2016 (last accessed July 25, 2019).

In the U.S., a recent study showed that 62% of adults in the U.S. get news on social media.[98] Furthermore, fake election news stories on Facebook were shared much more than the content posted by mainstream news outlets.[99] What is most unfortunate is the fact that many people reportedly believe these fake news stories.[100] Several commentators such as Hannah Jane Parkinson,[101] Max Read,[102] and Caitlin Dewey[103] have suggested that Donald Trump would not have been elected president were it not for the influence of fake news.

Aside from fake news, impersonations made on social media have also become rampant. Due to its fast-paced nature, "counterfeit Facebook and Twitter communications can damage the reputation of individuals and companies." It is because of this danger that a 2004 study[104] concluded that

98. Jeffrey Gottfried & Elisa Shearer, News Use Across Social Media Platforms 2016, *available at* http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016 (last accessed July 25, 2019).

99. Silverman, *supra* note 18.

100. Silverman & Singer-Vine, *supra* note 18.

101. Hannah Jane Parkinson, *Click and elect: how fake news helped Donald Trump win a real election*, THE GUARDIAN, Nov. 14, 2016, *available at* https://www.theguardian.com/commentisfree/2016/nov/14/fake-news-donald-trump-election-alt-right-social-media-tech-companies (last accessed July 25, 2019).

102. Max Read, Donald Trump Won Because of Facebook, *available at* http://nymag.com/selectall/2016/11/donald-trump-won-because-of-facebook.html (last accessed July 25, 2019).

103. Caitlin Dewey, *Facebook fake-news writer: 'I think Donald Trump is in the White House because of me'*, WASH. POST, Nov. 17 2016, *available at* https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me (last accessed July 25, 2019).

104. Monroe Price & Stefaan Verhulst, The Concept of Self-Regulation and the Internet (unpublished paper, University of Pennsylvania), *available at* https://repository.upenn.edu/cgi/viewcontent.cgi?article=1143&context=asc_papers (last accessed July 25, 2019).

the Internet has turned into quite a "scary place" because it has led people wanting to commit suicide due to inappropriate content spread through social media.

B. Understanding the Concept of Echo Chambers

Echo chambers are more commonly referred to as *bubbles*.[105] These refer to a group of users of a certain social media platform where they either intentionally or unintentionally see similar content with only very few alternative information. These echo chambers are created by reason of an algorithm of these social media companies.[106] It is normally based on one's online behavior.[107] A report made by the U.N. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression observed that social media platforms use algorithmic predictions of user preferences, which guide the advertisements individuals see, how their social media feeds are arranged, and the order in which search results appear.[108] The logic is quite simple in echo chamber discussion: if one is for a certain

105. *See generally* Seth Flaxman, et al., *Filter Bubbles, Echo Chambers, and Online News Consumption*, PUB. OPINION Q., Volume No. 80, Special Issue, at 299 (citing ELI PARISER, THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU (2011) & CASS R. SUNSTEIN, REPUBLIC.COM 2.0 (2009)) & Ivan Dylko, et al., *The dark side of technology: An experimental investigation of the influence of customizability technology on online political selective exposure*, 73 COMPUTERS IN HUMAN BEHAV. 181, 188 (2017) (citing Cass R. Sunstein, *The Law of Group Polarization*, 10 J. POL. PHIL. 175 (2002) & PARISER, *supra* note 106).

106. Dominic Spohr, *Fake news and ideological polarization: Filter bubbles and selective exposure on social media*, 34 BUS. INFO. REV. 150, 152-53 (2017) (citing PARISER, *supra* note 107, at 9).

107. *See* Christine Warner, This Is Exactly How Social Media Algorithms Work Today, *available at* https://www.skyword.com/contentstandard/marketing/this-is-exactly-how-social-media-algorithms-work-today (last accessed July 25, 2019).

108. Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, ¶ 21, 32d Session of the Human Rights Council, U.N. Doc. A/HRC/32/38 (May 11, 2016) (by David Kaye).

position, one will likely see more content favoring one's position. Clearly, the algorithm creating this echo chamber can have dangerous effects.

These *bubbles* are understandable pursuant to human psychology. People have a tendency to ignore facts that force their brains to work harder,[109] that is why social media users have no problem getting stuck in these echo chambers. By remaining trapped in these *bubbles*, their biases and beliefs are affirmed by like-minded users. This reality is affirmed by Facebook's research showing that the algorithm prioritizes "updates that users find *comforting*."[110] People nowadays prefer that what they think is confirmed by sources or information that share the same opinion.[111] In fact, even if one seeks to correct the culture of misinformation, it will not necessarily result in the change of one's belief.[112]

The danger posed by echo chambers is that it will greatly distort public opinion.[113] Society shall be greatly divided by reason of these *bubbles* because the public shall be fragmented. The more that people are trapped in these *bubbles*, the less people would think they share a reality with others outside the bubble. The moment that they lose this concept, it will seriously endanger democracy.[114]

---

109. *Yes, I'd Lie to You*, ECONOMIST, Sep. 10, 2016, *available at* https://www.economist.com/briefing/2016/09/10/yes-id-lie-to-you (last accessed July 25, 2019).

110. Zeynep Tufekci, *Mark Zuckerberg is in Denial*, N.Y. TIMES, Nov. 15, 2016, *available at* https://www.nytimes.com/2016/11/15/opinion/mark-zuckerberg-is-in-denial.html (last accessed July 25, 2019) (emphasis supplied).

111. David Lazer, et al., Combating Fake News: An Agenda for Research and Action, *available at* https://shorensteincenter.org/combating-fake-news-agenda-for-research (last accessed July 25, 2019).

112. *Id.* (citing Brenhan Nyhan & Jason Reifler, *When corrections fail: The persistence of political misperceptions*, 32 POL. BEHAVIOR 2, 303, 323 (2010) & D.J. Flynn, et al., *The Nature and Origins of Misconceptions: Understanding False and Unsupported Beliefs About Politics,* 38 ADVANCES IN POL. PSYCHOL. 127, 130 (2017)).

113. Flynn, et al., *supra* note 112, at 127 & 143.

114. *Id.* at 144.

C. Rise of the Trolls

Online trolls can also take the form of *bots* or special programs assigned to operate without need of any control from a user, and made to pretend to be a genuine user.[115] *Bots* are powerful enough to send content to a large number of users and share fake news at great speeds through the use of programming.[116] *Bots* can also like, friend, and follow each other, which results in a seemingly active and genuine profile.[117] The main purpose of these *bots* is to invade social media platforms with content and bombard them with these information.[118] They also act as trolls when they are used to harass people on social media in order to advocate a stand or position in line with the political machinery using them.[119] This is evident in 2015 when five percent of tweets were made by *bots*.[120] It does not help democracy when the ones speaking up are not even human beings in the first place.

III. SOCIAL MEDIA REGULATION

A. Challenges of Social Media Regulation

Regulating content shared through social media can be very challenging, especially because it poses serious issues against one's freedom of expression.[121]

The main question is, does this freedom extend to social media? It seems that the answer is in the affirmative.[122] Just like in other forms of

115. HEGELICH, *supra* note 90, at 2.

116. *Id.* at 3.

117. *Id.* at 5.

118. *See* Carl Miller, Governments don't set the political agenda anymore, bots do, *available at* http://www.wired.co.uk/article/politics-governments-bots-twitter (last accessed July 25, 2019).

119. Stephan Russ-Mohl, Research: When Robots Troll, *available at* https://en.ejo.ch/digital-news/spamming-robots (last accessed July 25, 2019).

120. *Id.*

121. *See* Wu, *supra* note 37, at 290.

122. *Id.*

communication, the freedom must be respected.[123] However, "[r]estrictions on freedom of expression on the Internet are only acceptable if they comply with established international standards, including that they are provided by law, and that they are necessary to protect an interest which is recognized under international law[.]"[124]

Social media's cross-border nature justifies the application of generally accepted principles of international law.[125] The freedom to express one's thoughts and opinions is primarily enshrined under Article 19 of the UDHR, which states, "[e]veryone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive[,] and impart information and ideas through any media and regardless of frontiers."[126] The UDHR is not a treaty or an international agreement. It is rather a declaration which announces to the world the main principles of human rights and freedoms "as a common standard of achievement for all peoples of all nations."[127]

The ICCPR is another source of the freedom. Aside from being a legally binding document,[128] it finds application in social media. A State party is therefore required to enforce the rights provided under Article 19 in the internet.[129] Under the ICCPR, "[e]veryone shall have the right to freedom of expression; this right shall include freedom to seek, receive[,] and impart

123. *Id.*

124. THE REPRESENTATIVE ON FREEDOM OF THE MEDIA ORGANIZATION FOR SECURITY AND CO-OPERATION IN EUROPE, JOINT DECLARATIONS OF THE REPRESENTATIVES OF INTERGOVERNMENTAL BODIES TO PROTECT FREE MEDIA AND EXPRESSION 66 (Adeline Hulin ed., 2013).

125. DRAGOS CUCEREANU, ASPECTS OF REGULATING FREEDOM OF EXPRESSION ON THE INTERNET 216 (2008).

126. UDHR, *supra* note 33, art. 19.

127. Hurst Hannum, *The Status of the Universal Declaration of Human Rights in National and International Law*, 25 GA. J. INT'L & COMP. L. 287, 289 (1996) (citing UDHR, *supra* note 33, pmbl.)

128. *See* CUCEREANU, *supra* note 125, at 219.

129. *See generally* Zhanna Kozhamberdiyeva, *Freedom of Expression on the Internet: A Case Study of Uzbekistan*, 33 REV. CENT. & EAST EUR. L. 95, 98–103 (2008).

information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."[130] The ICCPR binds all 173 parties and 74 signatories, including the Philippines, to abide by the rights and freedoms expressed therein.[131]

Aside from the generally accepted principles of international law, domestic laws are equally relevant and find application in social media. In the U.S., jurisprudence shows that the protection embodied under the First Amendment[132] extends to online speech.

In *Bland v. Roberts*,[133] the 4th U.S. Circuit Court of Appeals ruled that "[o]n the most basic level, clicking on the 'like' button literally causes to be published the statement that the [u]ser 'likes' something, which is itself a substantive statement."[134] The court ruled that a Facebook user who likes a Facebook page "engage[s] in legally protected speech[.]"[135]

Article 10 of the European Convention on Human Rights (ECHR) is likewise applicable to social media content made within the European Union

130. ICCPR, *supra* note 40, art. 19, ¶ 2.

131. *See* Status of Treaties: International Covenant on Civil and Political Rights (A Sub-chapter from the Online Depositary of the United Nations Treaty Collection) at 1, 2, & 40, *available at* https://treaties.un.org/doc/Publication/MTDSG/Volume%20I/Chapter%20IV/IV-4.en.pdf (last accessed July 25, 2019).

132. U.S. CONST. amend. I. The First Amendment provides that "Congress shall make no law ... abridging the freedom of speech, or of the press ..." U.S. CONST. amend. I.

133. Bland v. Roberts, 730 F.3d 368 (4th Cir. 2013) (U.S.).

134. *Id.* at 386.

135. Jonathan Stempel, Facebook 'like' deserves free speech protection: U.S. court, *available at* http://reut.rs/1dipEFg (last accessed July 25, 2019) (citing *Bland*, 730 F.3d at 386).

(E.U.)'s jurisdiction.[136] It protects the information regardless of the medium or channel of distribution.[137]

The Philippines is no exception. The right to free speech is expressly provided under Article III of the 1987 Constitution.[138] In several Supreme Court decisions, it has been ruled that the freedom extends even to cyberspace.[139]

While the right to free speech is oftentimes the general rule, it is subject to several exceptions. Under Article 29 (2) of the UDHR, it states that freedom of expression is subject "to such limitations as are determined by law solely for the purpose of securing due recognition and respect for the rights and freedom[ ] of others and of meeting the just requirements of morality, public order[,] and the general welfare in a democratic society."[140]

Even Article 19 (3) of the ICCPR provides that the freedom cannot be used to frustrate the rights of others.[141] The Convention allows the restriction of the right to freedom of expression for certain issues such as the "protection of national security" and "public order."[142]

Yet, for these limitations to take effect, it must be provided for by law.[143] It must also be justified by the State party by showing a certain degree of

136. *See* European Convention for the Protection of Human Rights and Fundamental Freedoms, as amended by Protocols Nos. 11 and 14 art. 10, ETS 5 [hereinafter ECHR].

137. Eva Lievans, Peggy Valke, David Stevens, & Pieter-Jan Ombelet, Freedom of Speech, Vanden Broele, (2015).

138. PHIL. CONST. art. III, § 4.

139. *See* Chavez v. Gonzales, 545 SCRA 441, 485 (2008) & *Disini, Jr.*, 716 SCRA at 344.

140. UDHR, *supra* note 33, art. 29, ¶ 2.

141. *See* ICCPR, *supra* note 40, art. 19, ¶ 3 (a).

142. ICCPR, *supra* note 40, art. 19, ¶ 3 (b).

143. *Id.* ¶ 3.

necessity.[144] It is very important that a legitimate goal must be answered for any restriction to be imposed.[145]

A good example of a valid limitation to freedom of expression is the 2001 Cybercrime Convention. It sought to regulate the production, offering, distribution, procuring, and possession of child pornography, as mentioned in Article 9 of the Convention,[146] and to punish racial comments or hate speech, including the denial of genocide.[147]

Other examples of valid limitations can be found in the U.S. and in the E.U., where their respective laws acknowledge the existence of certain exceptions to the freedom, such as the prevention of hate speech, defamation, or threats.[148]

According to U.S. jurisprudence, First Amendment rights extend even in cyberspace.[149] Nonetheless, there are cases which show that the right does not give others the right to defame another; hence, a party victim of defamation may seek redress from the courts.[150]

On the other hand, the ECHR also recognizes that the freedom is not absolute. It states that "[t]he exercise of these freedoms[ ] ... may be subject to

---

144. *Id.*

145. CUCEREANU, *supra* note 125, at 218.

146. Convention on Cybercrime art. 9, *opened for signature* Nov. 23, 2001, E.T.S. 185.

147. *See* 2003 Session of the Parliamentary Assembly, Strasbourg, France, Jan. 28, 2003, *Explanatory Report to the Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems*, ¶¶ 27, 33, & 40.

148. Marie-Andrée Weiss, Regulating Freedom of Speech on Social Media: Comparing the EU and the U.S. (An Abstract of a Research Project Published Online by Stanford Law School), *available at* https://law.stanford.edu/projects/regulating-freedom-of-speech-on-social-media-comparing-the-eu-and-the-u-s-approach (last accessed July 25, 2019).

149. Mobilisa, Inc. v. Doe, 170 P.3d 712, 717 (Ariz. Ct. App. 2007) (U.S.) (citing Reno v. American Civil Liberties Union, 521 U.S. 844, 870 (1997) (U.S.)).

150. Beauharnais v. Illinois, 343 U.S. 250, 266 (1952). The court provided that "[l]ibelous utterances [are] not... within the area of constitutionally protected speech[.]" *Id.*

such formalities, conditions, restrictions[,] or penalties as are prescribed by law and are necessary in a democratic society."[151]

The principle that this freedom is subject to limitation is recognized under Philippine law and jurisprudence. The Commission on Human Rights, a political entity empowered to ensure that human rights are protected, has expressed that one's "right to free speech has its [bounds], based on both international and domestic law."[152]

The Supreme Court of the Philippines gave emphasis on this matter by using cyberlibel as an example. In *Disini, Jr.*, the Court gave emphasis that cyberlibel is "not a constitutionally protected speech and that the government has an obligation to protect ... individuals from defamation."[153] In its initial discussion, it affirmed the government's duty to impose restrictions on social media — "For this reason, the government has a legitimate right to regulate the use of cyberspace and contain and punish wrongdoings."[154]

Nonetheless, it seems apparent that due to the advances of technology, existing laws, both domestic and international, are insufficient to protect the rights of innocent people from being damaged due to the proliferation of harmful content. Furthermore, society is greatly affected by the spread of fake news, hate speech, and the rise of internet trolls.[155]

Because of these instances, social networking sites have chosen to take it upon themselves to present a solution. Facebook, one of the biggest social media sites in the world, has taken note and has vowed to take action of the

151. ECHR, *supra* note 136, art. 10, ¶ 2.

152. Janvic Mateo, *Freedom of speech not absolute – CHR*, PHIL. STAR, May 27, 2016, *available at* http://www.philstar.com/headlines/2016/05/27/1587433/freedom-speech-not-absolute-chr (last accessed July 25, 2019).

153. *Disini, Jr.*, 716 SCRA at 320.

154. *Id.* at 298.

155. A troll is "a person who makes a deliberately offensive or provocative online post[.]" Lexico.com, Troll, *available at* https://www.lexico.com/en/definition/troll (last accessed July 25, 2019).

spread of fake news in its site.[156] Google, on the other hand, has also taken the lead in increasing regulation of false ads and fake news appearing on its system.[157]

It is because of these threats that States have considered taking initiative in lobbying for social media regulation. Germany recently approved, through its Cabinet, a bill that seeks to punish social networking sites if they fail to quickly remove content involving hate speech or defamatory fake news.[158] The bill seeks to impose heavy fines on social networking sites if they act slow in removing harmful content.[159]

B. Constitutionality of Social Media Regulation

Both social media and internet regulations, being relatively new modes of communication, are still constantly being challenged for their constitutionality in various States.

The Convention on Cybercrime is considered as a success story insofar as social media regulation is concerned because it gives rise to the possibility of having, at the very least, a limited treaty on social media. It is a fact that

> various states signed the Convention on Cybercrime. As a result, there is a plausible argument that the U.N. could pass a limited convention that only

156. Mark Molloy, *Facebook just made it harder for you to share fake news*, TELEGRAPH, Mar. 20, 2017, http://www.telegraph.co.uk/technology/2017/03/20/facebook-just-made-harder-share-fake-news (last accessed July 25, 2019).

157. Charles Warner, Google Increases Regulation of False Ads and Fake News, *available at* https://www.forbes.com/sites/charleswarner/2017/01/25/google-increases-regulation-of-false-ads-and-fake-news/#73533b4513f2 (last accessed July 25, 2019).

158. CNBC, Germany approves bill curbing online hate crime, fake news, *available at* http://www.cnbc.com/2017/04/06/germany-fake-news-fines-facebook-twitter.html (last accessed July 25, 2019).

159. Anthony Faiola & Stephanie Kirchner, *How do you stop fake news? In Germany, with a law*, WASH. POST, Apr. 5, 2017, *available at* https://www.washingtonpost.com/world/europe/how-do-you-stop-fake-news-in-germany-with-a-law/2017/04/05/e6834ad6-1a08-11e7-bcc2-7d1a0973e7b2_story.html?utm_term=.eb465edd752a (last accessed July 25, 2019).

> covers social media. However, cyberterrorism and social media have fundamental differences in the harms they create and the ability to identify violations.[160]

In both instances, cyberterrorism and social media pose harmful threats if abused.[161] "Terrorists have increasingly used social media as a recruiting [place] and publicity tool."[162] Thus, both require intensive monitoring because concerns arising from the spread of online content are alarming.[163] Hence, "regulation to limit and remove harmful content has the potential to save lives."[164] However, several gray areas present when identifying if social media content should be regulated is not the same in cyberterrorism, because in the latter, it is clear when the act is committed.[165]

In the U.S., the high regard of its courts for the First Amendment right has led to the declaration of unconstitutionality of various legislation seeking to regulate online content, specifically child pornography. In *American Civil Liberties Union v. Gonzales*,[166] the issues are "the constitutionality of the Child Online Protection Act, 47 U.S.C. § 231 (COPA) and whether [the] court should issue a permanent injunction against its enforcement due to its alleged constitutional infirmities."[167] In addition, "COPA provides both criminal and civil penalties for transmitting sexually explicit materials and communications over the World Wide Web ('Web') which are available to minors and harmful to them."[168] The court concluded that "COPA facially violates the First and Fifth Amendment rights ... ."[169] First, the "[d]efendant has failed to

160. Wu, *supra* note 37, at 284.

161. *Id.* at 298.

162. *Id.* at 283.

163. *Id.* at 298.

164. *Id.* at 283.

165. *Id.* at 299.

166. American Civil Liberties Union v. Gonzales, 478 F.Supp.2d 775 (E.D. Pa. 2007) (U.S.).

167. *Id.* at 777.

168. *Id.* (citing 47 U.S.C. § 231 (a) (West 2019) (U.S.)).

169. *Gonzales*, 478 F.Supp.2d at 777.

successfully defend against the plaintiffs' assertion that filter software and the Government's promotion and support thereof is a less restrictive alternative to COPA."[170]

> Filters are less restrictive than COPA. They impose selective restrictions on speech at the receiving end, not universal restrictions at the source. Under a filtering regime, adults without children may gain access to speech they have a right to see without having to identify themselves or provide their credit card information. Even adults with children may obtain access to the same speech on the same terms simply by turning off the filter on their home computers. Above all, promoting the use of filters does not condemn as criminal any category of speech, and so the potential chilling effect is eliminated, or at least much diminished. All of these things are true, moreover, regardless of how broadly or narrowly the definitions in COPA are construed.[171]

Also, the "[d]efendant has also failed to show that filters are not at least as effective as COPA at protecting minors from harmful material on the Web."[172] Second, COPA is vague.[173] "A party cannot bring a facial vagueness claim if the challenged regulation clearly applies to that party's speech."[174] The court said that in this case,

> Congress intended COPA to apply only to commercial pornographers ... However, the plaintiffs in this action are not commercial pornographers, a fact which has not escaped [the] defendant's notice in his challenges to their standing. Therefore, because COPA does not clearly apply to the plaintiffs' speech, the plaintiffs may bring a facial vagueness claim.[175]

---

170. *Id.* at 813.

171. *Id.* (citing Ashcroft v. American Civil Liberties Union, 542 U.S. 656, 667 (2004) (U.S.)).

172. *Gonzales*, 478 F.Supp.2d at 814.

173. *Id.* at 816.

174. *Id.* (citing Gibso v. Mayor and Council of City of Wilmington, 355 F.3d 215, 225 (3d Cir. 2004) (U.S.) & Rode v. Dellarciprete, 845 F.2d 1195, 1200 (3d Cir. 1988) (U.S.)).

175. *Gonzales*, 478 F.Supp.2d at 816.

Lastly, COPA is overbroad.[176] "The overbreadth doctrine prohibits the Government from banning unprotected speech if a substantial amount of protected speech is prohibited or chilled in the process."[177] The court explains,

> Since the vagueness of 'communication for commercial purposes' and 'engaged in business' would allow prosecutors to use COPA against not only Web publishers with commercial Web sites who seek profit as their primary objective but also those Web publishers who receive revenue through advertising or indirectly in some other manner, the array of Web sites to which COPA could be applied is quite extensive. Such a widespread application of COPA would prohibit and undoubtedly chill a substantial amount of constitutionally protected speech for adults.[178]

Another U.S. case pertinent to constitutionality of any regulation to freedom of speech or expression is *Reno v. American Civil Liberties Union*.[179] In 1997, the United States Supreme Court struck down the Communications Decency Act as unconstitutional.[180] In this case, the law sought to regulate obscenity and indecency of children in cyberspace.[181] The U.S. Supreme Court held that the Act violated the First Amendment because its regulations amounted to a content-based blanket restriction of free speech.[182] The Act failed to clearly define *indecent* communications, limit its restrictions to particular times or individuals (by showing that it would not impact adults), provide supportive statements from an authority on the unique nature of internet communications, or conclusively demonstrate that the transmission of *offensive* material is devoid of any social value.[183] The court added that since the First Amendment distinguishes between *indecent* and *obscene* sexual

176. *Id.* at 819.

177. *Id.* (citing Ashcroft v. Free Speech Coalition, 535 U.S. 234, 255 (2002) (U.S.)).

178. *Gonzales*, 478 F.Supp.2d at 819.

179. Reno v. American Civil Liberties Union, 521 U.S. 844 (1997) (U.S.).

180. *Id.* at 874 & 882.

181. *Id.* at 859-60.

182. *Id.* at 871-72 & 879.

183. *Id.* at 875, 867, & 881.

expressions, protecting only the former, the Act could be saved from facial overbreadth challenges if the words "or indecent" are severed from its text.[184]

In the E.U. the freedom has been recently challenged in relation to an individual's right to privacy. In the landmark case of *Google Spain SL and Google, Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja Gonzalez*,[185] the court laid down a clear indication that the freedom of expression on social media has its limitations, and it can be regulated by deleting certain types of content,[186] otherwise known as the *right to be forgotten*.[187] The E.U. Court said that individuals have the right — under certain conditions — to ask search engines to remove links with personal information about them.[188] This applies where the information is inaccurate, "inadequate, irrelevant ... or excessive" for the purposes of the data processing.[189] The court found that in this particular case the interference with a person's right to data protection could not be justified merely by the economic interest of the search engine.[190] At the same time, the court explicitly clarified that the right to be forgotten is not absolute but will always need to be balanced against other fundamental rights, such as the freedom of expression and of the media.[191] A case-by-case assessment is needed considering the type of information in question, its sensitivity for the individual's private life, and the interest of the public in having access to that

184. *Id.* at 883.

185. Google Spain SL and Google Inc. v. Agencia Española de Protección de Datos (AEPD) and Mario Costeja González, Judgment, Case C-131/12, EU:C:2014:317 (CJEU May 13, 2014).

186. *Id.* ¶ 99.

187. *Id.* ¶ 91.

188. *Id.* ¶ 96.

189. *Id.* ¶ 94.

190. *Id.* ¶ 81.

191. *Google Spain SL and Google Inc.*, EU:C:2014:317, ¶ 81.

information.[192] The role that the person requesting the deletion plays in public life might also be relevant.[193]

Also, in the United Kingdom, the Communications Act of 2003, one of the first statutes to regulate online speech, has led to several convictions.[194] One of which was a certain Darryl O'Donnell who posted messages on Facebook saying that someone was "a 'scumbag' and should 'get a bullet in the head.'" [195] The district judge said that "O'Donnell's comments were menacing and offensive and should not have been posted on Facebook."[196]

Insofar as online impersonations or "Internet trolls" are concerned, the U.S. Supreme Court has laid down guidelines to ensure that defamation plaintiffs may seek redress and find out who are behind the accounts spreading harmful content.

---

192. *Id.*

193. *Id. See also* European Commission, Factsheet on the "Right to be Forgotten" ruling, *available at* https://www.inforights.im/media/1186/cl_eu_commission_factsheet_right_to_be-forgotten.pdf (last accessed July 25, 2019).

194. See An Act to confer functions on the Office of Communications; to make provision about the regulation of the provision of electronic communications networks and services and of the use of the electro-magnetic spectrum; to make provision about the regulation of broadcasting and of the provision of television and radio services; to make provision about mergers involving newspaper and other media enterprises and, in that connection, to amend the Enterprise Act 2002; and for connected purposes [Communications Act 2003], 2003 c. 21 (2003) (U.K.).

195. Ian Cram, The Protection of Human Rights in the UK Constitution: Freedom of Expression and Social Media (Conference Paper Presented at National Taipei University) at *8, *available at* https://www.biicl.org/documents/550_taiwan_uk_project_-_prof_ian_cram_conference_report_final_28_4_2015.pdf?showdocument=1 (last accessed July 25, 2019).

196. BBC News, Man fined for Gregory Campbell Facebook comment, *available at* http://www.bbc.com/news/uk-northern-ireland-14345649 (last accessed July 25, 2019).

In *Doe v. Cahill*,[197] which modified *Dendrite Intern., Inc. v. Doe No. 3*,[198] the court has laid down the *Dendrite-Cahill* standard where a plaintiff must provide sufficient notice to anonymous posters that they are the subject of an application to disclose their identity; identify the exact statements which are actionable; and provide the court with sufficient evidence to establish a *prima facie* case.[199] After doing so, the court will then balance the defendant's right under the First Amendment against the strength of the evidence presented.[200]

Furthermore, in *USA Technologies, Inc, v. Doe*,[201] the Court of the Northern District of California quashed the subpoena seeking for the reveal of the online poster because the evidence presented was insufficient to disfavor the defendant's First Amendment right, considering that the statements pointed out were deemed as "rhetorical hyperbole."[202] The court said that "statements which are merely annoying or embarrassing or no more than rhetorical hyperbole or a vigorous epithet are not defamatory."[203] In this case, the statement complained of asserts that "[Jensen is] a caricature of any number of characters in Dickens or Shakespeare whose worldview is that humanity exists to be fleeced."[204] Although this statement seems offensive, defamation "does not extend to mere insult."[205]

Nevertheless, there are courts that have already granted a plaintiff's petition to have the poster's identity subject to a subpoena. In *Maxon v. Ottawa Publishing*,[206] the Appellate Court of Illinois ruled in favor of the plaintiff

197. Doe v. Cahill, 884 A.2d 451 (Del. 2005) (U.S.).

198. Dendrite Intern., Inc. v. Doe No. 3, 775 A.2d 756 (N.J. Super. Ct. App. Div. 2001) (U.S.).

199. *Doe*, 884 A.2d at 460 (citing *Dendrite Intern., Inc.*, 775 A.2d at 760).

200. *Doe*, 884 A.2d at 460 (citing *Dendrite Intern., Inc.*, 775 A.2d at 760-61).

201. USA Technologies, Inc. v. Doe, 713 F.Supp.2d 901 (N.D. Cal. 2010) (U.S.).

202. *Id.* at 908.

203. *Id.* (citing Beverly Enterprises, Inc. v. Trump, 182 F.3d 183, 187 (3d Cir. 1999) (U.S.)).

204. *USA Technologies, Inc.*, 713 F.Supp.2d at 908.

205. *Id.* (citing *Beverly Enterprises, Inc.*, 182 F.3d at 187).

206. Maxon v. Ottawa Pub. Co., 929 N.E.2d 666 (Ill. App. Ct. 3d Dist. 2010) (U.S.).

because the statements complained of were actionable.[207] The court said in disclosing the identity of any anonymous potential defamation defendant,

> the court must insure that the petitioner: (1) is verified; (2) states with particularity facts that would establish a cause of action for defamation; (3) seeks only the identity of the potential defendant and no other information necessary to establish the cause of action of defamation; and (4) is subjected to a hearing at which the court determines that the petition sufficiently states a cause of action for defamation against the unnamed potential defendant[.][208]

In the Philippines, the Supreme Court has yet to encounter an opportunity to rule on the issue of *Internet trolls*. The closest the Court has been in ruling on matters involving the cyberspace is in *Disini, Jr.*[209] In this case, the Supreme Court discussed the overbreadth doctrine, which states that "a proper governmental purpose, constitutionally subject to state regulation, may not be achieved by means that unnecessarily sweep its subject broadly, thereby invading the area of protected freedoms."[210] Thus, it would be a challenge for the Philippine government to justify the regulation of the so-called *Internet trolls* since the definition could be so broad where it could violate one's freedom of speech or expression.

As mentioned earlier, freedom of expression is not absolute. It is subject to certain limitations that are enforceable through the police power of the state. In *ABS-CBN Broadcasting Corp. v. Commission on Elections*,[211] the Supreme Court ruled that freedoms of speech, of expression, and of the press are "not immune to regulation by the State in the exercise of its police power."[212] In ruling against the COMELEC resolution and stating that

207. *Id.* at 675-76.

208. *Id.* at 673.

209. *See Disini Jr.*, 716 SCRA.

210. *Disini, Jr.*, 716 SCRA at 303 (citing Southern Hemisphere Engagement Network, Inc. v. Anti-Terrorism Council, 632 SCRA 146, 185 (2010)).

211. ABS-CBN Broadcasting Corp. v. Commission on Elections, 323 SCRA 811 (2000).

212. *Id.* at 824 (citing Badoy, Jr. v. Comelec, 35 SCRA 285, 289 (1970)).

prohibiting exit polls does violate the freedom of expression clause enshrined in the Constitution, the Supreme Court said that

> [t]he interest of the [State] in reducing disruption is outweighed by the drastic abridgment of the constitutionally guaranteed rights of the media and the electorate. Quite the contrary, instead of disrupting elections, exit polls [—] properly conducted and publicized [—] can be vital tools for the holding of honest, orderly, peaceful[,] and credible elections[,] and for the elimination of election-fixing, fraud[,] and other electoral ills.[213]

Also, as early as 1912, the Supreme Court said that

> [t]he enjoyment of a private reputation is as much a constitutional right as the possession of life, liberty[,] or property. It is one of those rights necessary to human society that underlie the whole scheme of human civilization. The law recognizes the value of such a reputation, and constantly strives to give redress for its injury. It imposes upon him [or her] who attacks it by slanderous words, or libelous publication, a liability to make full compensation for the damage to the reputation, for the shame and obloquy, and for the injury to the feelings of the owner, which are caused by the publication of the slander or the libel.[214]

One of the existing regulations to one's freedom of expression in the Philippines is through the crime of libel under the Revised Penal Code.[215] It defined libel as "a public and malicious imputation of a crime, or of a vice or defect, real or imaginary, or any act, omission, condition, status, or circumstance tending to cause the dishonor, discredit, or contempt of a natural or juridical person, or to blacken the memory of one who is dead."[216]

Furthermore, in *Lopez v. People*,[217] the Supreme Court held that libel is among the exceptions to one's right to free speech.[218] The High Court said that

---

213. *ABS-CBN Broadcasting Corp.*, 323 SCRA at 830.

214. *Worcester*, 22 Phil. at 98.

215. An Act Revising the Penal Code and Other Penal Laws [REVISED PENAL CODE], Act No. 3815, art. 353 (1930).

216. *Id.*

217. Lopez v. People, 642 SCRA 668 (2011).

218. *Id.* at 671.

> [f]reedom of expression enjoys an exalted place in the hierarchy of constitutional rights. Free expression however, 'is not absolute for it may be so regulated; that its exercise shall neither be injurious to the equal enjoyment of others having equal rights, nor injurious to the rights of the community or society.'[219]

Again, in *Disini, Jr.*, the Supreme Court defined cyberspace as "a system that accommodates millions and billions of simultaneous and ongoing individual accesses to and uses of the internet."[220]

> But all is not well with the said system since it could not filter out a number of persons of ill will who would want to use cyberspace technology for mischiefs and crimes. One of them can, for instance, avail himself of the system to unjustly ruin the reputation of another or bully the latter by posting defamatory statements against him that people can read.[221]

Thus, the Philippine Supreme Court ruled that the general objective of the Cybercrime Prevention Act is to "reasonably put order into cyberspace activities, punish wrongdoings, and prevent hurtful attacks on the system."[222] Basically, it seeks to regulate access to and use of cyberspace.[223]

In the said case, the Court gave emphasis that indeed, the cybercrime law provides a chilling effect.[224] However, this form of regulation is necessary, otherwise, "to prevent the State from legislating criminal laws because they instill such kind of fear is to render the [State] powerless in addressing and penalizing socially harmful conduct."[225]

The High Court also gave reference to the UDHR and ICCPR stating that "although everyone should enjoy freedom of expression, its exercise carries with it special duties and responsibilities. Free speech is not absolute. It

---

219. *Id.* (citing Primicias v. Fugoso, 80 Phil. 71, 75 (1948)).

220. *Disini, Jr.*, 716 SCRA at 298.

221. *Id.*

222. *Id.* at 299.

223. *Id.* at 297.

224. *Id.* at 304.

225. *Id.* (citing *Southern Hemisphere Engagement Network, Inc.*, 632 SCRA at 186 & Estrada v. Sandiganbayan, 369 SCRA 394, 441 (2001)).

is subject to certain restrictions, as may be necessary and as may be provided by law."[226]

However, the Court mentioned that certain provisions of the law, specifically Section 5, suffers from overbreadth because it broadly oversweeps the constitutionally guaranteed freedoms of the people.[227] The Court said that "[u]nless the legislature crafts a cyber libel law that takes into account its unique circumstances and culture, such law will tend to create a chilling effect on the millions that use this new medium of communication in violation of their constitutionally-guaranteed right to freedom of expression."[228]

IV. EXISTING REGULATORY FRAMEWORKS

A. The International Regulatory Framework for Social Media

According to *The Economist*, "[s]ocial media in Western countries operate in a specific environment of 'legal exceptionalism.'"[229] With very few exceptions, "companies are not [generally] responsible for the content published on their platforms." [230] This mentality originated twenty years ago when countries like the U.S. has made clear their position that these industries must be protected due to the "apparent lack of understanding of the potential of social media platforms."[231] By looking back at U.S. history, it is clear that the thrust of the U.S. government in amending the Communications Decency Act and the enactment of the Digital Millennium Copyright Act was made in line with

226. *Disini, Jr.*, 716 SCRA at 320 (citing ICCPR, *supra* note 40, art. 19, ¶¶ 2-3).

227. *Disini, Jr.*, 716 SCRA at 327.

228. *Id.* at 325.

229. Konrad Niklewicz, Weeding Out Fake News: An Approach to Social Media Regulation (A Paper Published by Wilfried Martens Centre for European Studies) at 29, *available at* https://issuu.com/centreforeuropeanstudies/docs/ces-weeding_out_fake_news_v3_web (last accessed July 25, 2019) (citing *Eroding Exceptionalism: Internet firms' legal immunity is under threat*, ECONOMIST, Feb. 11, 2017, *available at* https://www.economist.com/business/2017/02/11/internet-firms-legal-immunity-is-under-threat (last accessed July 25, 2019)).

230. Niklewicz, *supra* note 229, at 29.

231. *Id.*

this mentality.[232] There were very limited exceptions as to when these contents call for regulation.[233]

In 2000, the E.U. posited a similar stand when the E-Commerce Directive referred to social media as "intermediary service providers."[234] Through this classification, social media companies were deemed not to be liable for content in its platform as long as they did not have knowledge of the illegality of the content and that they took only a passive position in the process.[235] This means that social media companies are not liable, provided that they act simply as a medium where users can store and transmit content. However, the moment that what was stored in their platform is patently illegal content such as child pornography, the law allows the information to be immediately removed upon notification.[236] This kind of feature is more commonly known as the *notice and take down* procedure.[237]

In the E.U., the formulation of the implementing rules was left to the national legislating bodies.[238] Under these rules, it is the State, through its assigned government agency, that is tasked to notify the social media company of the harmful content which must be taken down. It is worth noting that social media companies do not have an obligation to look and search for these harmful contents, absent any notification.[239] Under the E-Commerce Directive, these social media companies have no duty to monitor the content passing through and stored in their platforms.[240]

---

232. *Id.*

233. *Id.*

234. Directive 2000/31/EC, of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce, in the Internal Market, 2000 O.J. (L 178) 1, 12 [hereinafter E.U. E-Commerce Directive].

235. *Id.* at 12–13.

236. *Id.* at 6 & 13.

237. *Id.* at 15.

238. *Id.* at 6.

239. *Id.* at 13.

240. E.U. E-Commerce Directive, *supra* note 234, at 13.

More recently, the European Commission has taken the initiative to adopt a code of conduct for these providers with the objective to curb harmful content on social media.[241] It also served as a way to improve the *notice and take down* procedure.[242] Social media companies that took part committed to remove illegal content covered by the code within 24 hours from notification.[243] Lastly, the code required the companies to be the ones to set their own guidelines as to what is harmful content.[244]

A question arises: Is the publication or dissemination of fake news considered *harmful content*? It seems that in the case of Anas Modamani, a 19-year-old Syrian refugee, it was not.[245] Modamani had a picture taken with German Chancellor Angela Merkel that made rounds on the various front pages in Germany.[246] The same picture, however, was posted on social media which linked Anas Modamani to the terrorist attacks in Berlin and Brussels — clearly, fake news.[247] Still, the fake news generated heavy social media traffic, with more than 32,000 shares, reactions, and comments.[248] Although there were attempts to debunk the story through a legitimate fact-based article,[249] it only generated less than half of the traffic gained by the assailed content. Modamani sought for an order from the Würzburg court for Facebook to stop

241. Press Release *by* the European Commission, *European Commission and IT Companies announce Code of Conduct on illegal online hate speech* (May 31, 2016) (on file with Author).

242. *Id*.

243. *Id*.

244. *Id*.

245. Niklewicz, *supra* note 229, at 30.

246. *Id*.

247. *Id*.

248. *Id*.

249. Alberto Nardelli & Craig Silverman, Hyperpartisan Sites and Facebook Pages are Publishing False Stories And Conspiracy Theories About Angela Merkel, *available at* https://www.buzzfeed.com/albertonardelli/hyperpartisan-sites-and-facebook-pages-are-publishing-false (last accessed July 25, 2019).

and disallow the reposting of the picture in its platform.[250] The court, however, disallowed his application on the ground that no law exists allowing such prayer to be granted.[251] Based on the ruling of the Würzburg court denying the application for injunction, it is clear that fake news, although harmful, remains protected until a law is passed subjecting it to state regulation.[252]

There have been other attempts to combat fake news such as in the case of Eva Glawischnig, Austria's Green Party leader.[253] She filed a complaint praying for the deletion of a fake news item involving her that was posted on Facebook.[254] In May 2017, the Vienna court ordered Facebook to take down the assailed content not only in Austria but in the entire world.[255] This effectively resulted in a situation where Facebook must prevent users, even in places where the European hate speech law is not implemented, from seeing the content.[256]

---

250. Philip Oltermann, *Syrian who took Merkel selfie sues Facebook over 'defamatory' posts*, GUARDIAN, Jan. 12 2017, *available at* https://www.theguardian.com/world/2017/jan/12/syrian-who-took-merkel-selfie-sues-facebook-over-defamatory-posts (last accessed July 25, 2019).

251. Melissa Eddy, *Selfie with Merkel by Refugee Became a Legal Case, but Facebook Won in German Court*, N.Y. TIMES, Mar. 7, 2017, *available at* https://www.nytimes.com/2017/03/07/business/germany-facebook-refugee-selfie-merkel.html (last accessed July 25, 2019).

252. *See* Eddy, *supra* note 251.

253. Derek Scally, *Fakes Brought to Book: Austrian Greens Take on Facebook*, IRISH TIMES, Dec. 15, 2016, *available at* https://www.irishtimes.com/business/technology/fakes-brought-to-book-austrian-greens-take-on-facebook-1.2906277 (last accessed July 25, 2019).

254. *Id.*

255. Shadia Nasralla, Austrian court rules Facebook must delete 'hate postings', Reuters, *available at* https://www.reuters.com/article/us-facebook-austria/austrian-court-rules-facebook-must-delete-hate-postings-idUSKBN1841IF (last accessed July 25, 2019).

256. *Id.*

The landmark case of *Google Spain SL and Google, Inc. v. AEPD and Mario Costeja Gonzalez* is also worth mentioning because of the *right to be forgotten* on Google as the European Court ruled in favor of the delinking of the content from search results made on Google's platform pursuant to the balancing of interests of the person involved and the right of the user to information.[257] The court noted that in this instance, free speech is not violated because the content is not actually deleted, rather, it is simply *de-linked*.[258]

Taking all of these into consideration, it can be observed that foreign courts have yet to establish clear standards as to when fake news can be properly subject to regulation. If one chooses to follow the Würzburg court's ruling, then a law must be passed prior to regulation. Perhaps, this is more sensible considering that the State must prescribe standards as to what is considered fake news. However, in light of the Vienna court's ruling, even in the absence of a fake news law, the court ordered the taking down of the fake news. Obviously, the approach differs per State, but the common denominator in these mentioned instances is that the social media company plays a very important role in both the publication and the taking down aspect of any harmful content. This will be further discussed in the subsequent Section of this Chapter.

B. The Philippines' Regulatory Framework for Social Media

In the landmark case of *Disini, Jr.*, the Supreme Court upheld the application of the libel provision under the Cybercrime Law on the internet, effectively to the country's existing laws to cyberspace.[259] The High Court highlighted the distinction of the offense saying that "cyberlibel brings with it certain intricacies, unheard of when the penal code provisions on libel were enacted. The culture associated with internet media is distinct from that of print."[260]

The *Disini, Jr.* ruling is the first of its kind insofar as the regulation of speech made on social media. Based on the Court's ruling, it merely treated

257. *Google Spain SL and Google Inc.*, EU:C:2014:317, ¶¶ 81 & 99.

258. Abraham L. Newman, What the "right to be forgotten" means for privacy in a digital age, *available at* http://science.sciencemag.org/content/347/6221/507.full (last accessed July 25, 2019).

259. *Disini, Jr.*, 716 SCRA at 320.

260. *Id.*

social media simply as another mode of committing the crime of libel. In fact, the Court made emphasis that Article 355 of the Revised Penal Code talks about libel being committed by means of writing or similar means, with the latter being construed to include social media.[261]

Following this interpretation, one can argue that other crimes provided under the Revised Penal Code, specifically Article 154, can be prosecuted by the State even if committed on social media.[262] Article 154 involves the felony of unlawful use of means of publication and unlawful utterances,[263] to wit —

> *Art. 154. Unlawful use of means of publication and unlawful utterances.* — The penalty of *arresto mayor* and a fine ranging from ₱200 to ₱1,000 pesos shall be imposed upon:
>
> Any person who by means of printing, lithography, or any other means of publication shall publish or cause to be published as news any false news which may endanger the public order, or cause damage to the interest or credit of the State.[264]

Unfortunately, there is no jurisprudence discussing Article 154 of the Revised Penal Code. Hence, issues of its application, or even the constitutionality of the article itself, is still subject to debate.

An affected party may also opt to file an action for damages under Article 2176 of the Civil Code.[265] Although the article may seemingly limit the award of moral damages to either libel or slander, Article 26 of the Civil Code can serve as additional basis to warrant damages arising from a fake news article published on social media.[266] To be a valid cause of action, the post must tend to "pry ... to the privacy [and peace of mind of another,]"[267] "meddl[e] ... or

---

261. *See Disini, Jr.*, 716 SCRA at 316.

262. *See* REVISED PENAL CODE, art. 154.

263. REVISED PENAL CODE, art. 154.

264. *Id.*

265. An Act to Ordain and Institute the Civil Code of the Philippines [CIVIL CODE], Republic Act No. 386, art. 2176 (1949).

266. *Id.* art. 26.

267. *Id.* art. 26 (1).

disturb[ ] the private life or family relations of another[,]"[268] "intrigue to cause another to be alienated from [his or her] friends"[269] or "vex[ ] or humiliate[e] another on account of his [or her] religious belief[ ], lowly station in life, place of birth, physical defect, or other personal condition."[270]

Despite these laws, existing social realities have encouraged the country's legislators to examine the fake news problem in the Philippines in the hopes of passing a law that can effectively solve the problem. In October 2017, the Senate Committee on Public Information and Mass Media commenced an inquiry on the massive proliferation of fake news in the Philippines.[271] Headed by Senator Grace Poe, the Committee conducted inquiries with many of the resource persons in agreement that there seems to be no need for any further legislation.[272] What is needed, according to them, is stronger enforcement and implementation of the laws, as well as heightened e-literacy campaigns to educate the public on the impact of fake news and digital responsibility.[273] In fact, both the President Communications Operations Office and investigative news organization Vera Files agree that social media regulation is not the solution.[274] Vera Files President Ellen Tordesillas exclaimed that any form of regulation "might infringe on press freedom"[275] and that it might be "a cure ... [far] worse than the disease."[276]

---

268. *Id.* art. 26 (2).

269. *Id.* art. 26 (3).

270. *Id.* art. 26 (4).

271. Senate tackles spread of 'fake news', *available at* http://nine.cnnphilippines.com/news/2017/10/04/Senate-spread-of-fake-news.html (last accessed July 25, 2019).

272. *Id.*

273. Regine Cabato, PCOO, Vera Files agree: No to social media regulation, yes to media literacy, *available at* https://cnnphilippines.com/news/2018/02/01/PCOO-Vera-Files-no-social-media-regulation-yes-media-literacy.html (last accessed July 25, 2019).

274. *Id.*

275. *Id.*

276. *Id.*

Professor Antonio G.M. La Viña, who was also a resource person during the investigations, posited an alternative to social media regulation.[277] Instead of focusing on the user who is the author of the content, he argues that it is the social media company who should be made accountable.[278] When asked by Sen. Francis "Kiko" Pangilinan if that kind of measure would violate free speech, he said:

> Well, I [do not] think [that is] violative of the right to free speech because ... the person who is speaking can still speak. But what [you are] stopping is the means of propagating it. In traditional media, they also do that — right? They also stop fake news through verification.[279]

Professor Florin Hilbay's suggestion was also enlightening. He focused on public officials who publish and disseminate fake news and whether they should be subject to some form of liability.[280] He argued that whenever a public officer publishes content on social media, there is a *badge of truth* which accompanies it.[281] He further comments that public officers are expected to be placed on a higher standard of accountability pursuant to the Constitution.[282] He suggests that a government agency should be created which shall serve as a hub where individuals can complain about the fake news published by a public officer, which in turn, that agency will determine the appropriate penalty to be imposed upon the public officer after proper investigation and resolution.[283]

---

277. Rappler, Video, *Part 2: Senate hearing on fake news online, 4 October 2017*, Oct. 5, 2017, YOUTUBE, *available at* https://www.youtube.com/watch?v=QjQkXWESUQw (last accessed July 25, 2019) [hereinafter Senate hearing on fake news online] (video from 1:16:22 to 1:18:04).

278. *Id.*

279. *Id.*

280. *See* Senate hearing on fake news online, *supra* note 277 (video from 2:17:05 to 2:24:18).

281. *Id.* (video from 2:17:05 to 2:20:25).

282. *Id.*

283. *Id.* (video from 2:20:26 to 2:24:18).

C. Voluntary Self-Regulating Framework of Social Media Companies

By the end of 2016, there have been numerous criticisms against social media companies regarding their apparent indifference to fake news. There are arguments opining that social media companies must accept that they are liable, to a certain extent, for the widespread proliferation of fake news in the world and that they have an obligation to act and prevent further problems caused by the situation.[284]

Social media companies heeded to these calls and quickly addressed the issue. Facebook made an announcement in December 2016 that it had adopted several measures to solve the problem.[285] The company emphasized that the problem is deeply rooted with spammers who spread fake news for financial gain.[286] Facebook decided that it is the users who can help with this issue by reporting the fake news circulated on the platform.[287] By simply clicking an option added by Facebook in its interface, the assailed content will then be assessed by an independent group of fact-checkers.[288] Should the content be false, it shall be flagged.[289] The company clarified that while the content can

284. Niam Yaraghi, How should social media platforms combat misinformation and hate speech?, *available at* https://www.brookings.edu/blog/techtank/2019/04/09/how-should-social-media-platforms-combat-misinformation-and-hate-speech (last accessed July 25, 2019). *See also* The New York Times Editorial Board, *Facebook and the Digital Virus Called Fake News*, N.Y. TIMES, Nov. 19, 2016, *available at* https://www.nytimes.com/2016/11/20/opinion/sunday/facebook-and-the-digital-virus-called-fake-news.html (last accessed July 25, 2019).

285. Adam Mosseri, Addressing Hoaxes and Fake News, *available at* https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news (last accessed July 25, 2019).

286. *Id.*

287. *Id.*

288. *Id.*

289. *Id.*

still be shared, other users will be warned about the dispute.[290] This was first launched in the U.S..[291]

Adam Mosseri, Facebook's Vice President, said: "We believe in giving people a voice," emphasizing that the social media company was not to act as a censor or even an arbiter of truth.[292]

In January of the following year, Facebook announced that these tools would soon be available in Germany, with Correctiv as the initial choice as the third-party, independent fact checker. [293] Unfortunately, Facebook's mechanism was put on hold for failing to find more partners.[294] Axel Springer, through its Chief Executive Officer Mathias Dôpfner, exclaimed its non-participation of this initiative by Facebook by saying that it would be wrong for publishers like itself to help social media companies solve their problem involving credibility.[295]

290. Alex Heath, Facebook is going to use Snopes and other fact-checkers to combat and bury 'fake news', *available at* https://www.businessinsider.com/facebook-will-fact-check-label-fake-news-in-news-feed-2016-12 (last accessed July 25, 2019).

291. Elle Hunt, *'Disputed by Multiple Fact-Checkers': Facebook Rolls Out New Alert to Combat Fake-News*, GUARDIAN, Mar. 22, 2017, *available at* https://www.theguardian.com/technology/2017/mar/22/facebook-fact-checking-tool-fake-news (last accessed July 25, 2019).

292. Mosseri, *supra* note 285.

293. Allen Cone, Facebook to Begin Labelling Fake News for German Users, *available at* https://www.upi.com/Facebook-to-begin-labeling-fake-news-for-German-users/4531484495660 (last accessed July 25, 2019).

294. Fabian Reinbold, Facebook, Fake News and the media: Enlightenment desperately wanted, *available at* http://www.spiegel.de/netzwelt/netzpolitik/fake-news-facebook-und-focus-online-stehen-vor-partnerschaft-a-1135013.html (last accessed July 25, 2019) (translate the webpage to English by clicking "Translate this page" button in the Google search results).

295. Guy Chazan, *Axel Springer chief rules out helping Facebook detect fake news*, FIN. TIMES, Apr. 7, 2017, *available at* https://www.ft.com/content/36fc025e-04bb-11e7-ace0-1ce02efodef9 (last accessed July 25, 2019).

The Netherlands and France were also chosen as areas of experimentation for these measures.[296] Motivated by the fact that these countries were then holding elections, the approach of Facebook, which was to rely on user engagement in figuring out which content was fake, was highly lauded by commentators saying that the best way to police social media is still through the users themselves.[297]

In April 2017, Facebook published a paper on *Information Operations*, wherein it emphasized its intention to fight fake news and illegal content published on their platform.[298] The paper's added value, which was co-authored by the company's senior security officers, is that it details the company's view on the definition of fake news and related problems.[299] For example, Facebook distinguishes *disinformation* as "inaccurate ... content ... spread intentionally"[300] from *misinformation* or the "unintentional spread of inaccurate information ..."[301] It also provides a definition for *false news* as items that "purport to be factual, but which contain intentional misstatements ... with the intention to arouse passion, attract viewership, or deceive."[302] Furthermore, Facebook has openly acknowledged the existence of *Information Operations* — actions taken by organized actors, both government and non-state actors, with the purpose of distorting the truth and political opinions, to

296. Madhumita Murgia, Facebook to pay fact-checkers to combat fake news, FIN. TIMES, Apr. 7, 2017, *available at* https://www.ft.com/content/ba7d4020-1ad7-11e7-a266-12672483791a (last accessed July 25, 2019).

297. Krzystof Iszkowski, Limits of freedom of speech in social media, *available at* http://wethecrowd.liberte.pl/granice-wolnosci-slowa-w-social-media (last accessed July 25, 2019) (translate web page to English by clicking "Translate this page" button in Google search results).

298. *See* Jen Weedon, et al., Information Operations and Facebook (A Paper Published by Facebook) at 3, *available at* https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf (last accessed July 25, 2019).

299. Weedon, et al., *supra* note 298, at 5.

300. *Id.*

301. *Id.*

302. *Id.*

strategically accomplish a certain agenda.[303] According to Facebook's research, there are three common objectives of *Information Operations*, namely: (1) "[p]romoting or denigrating a specific issue[;]"[304] (2) "[s]owing distrust in political institutions[;]"[305] and (3) "[s]preading confusion[.]"[306] Facebook's paper further acknowledged the reality that fake news can be amplified not only by the creators of disinformation through their own network of false accounts, but by everyday users as well, which results in authentic networks.[307] The company recognizes that the motivation of the false amplifiers is "ideological rather than financial."[308]

It also made additional recommendations as to how Facebook aims to fight malicious *Information Operations*. For example, it underlined the company's commitment to prevent and delete fake accounts, whether they are manually or automatically operated.[309] An example of this is the fact that during the France elections, Facebook is said to have suppressed 30,000 fake accounts.[310]

Social media companies, other than Facebook, have employed other interesting voluntary self-regulating measures to fight fake news on their platforms. Snap, the owner of the Snapchat platform, has asked third parties publishing on its *Discover* platform to vouch for the content they provide.[311] As the company's spokesperson explained, it wants its editorial partners "to do their part to keep Snapchat an informative, factual, and safe environment."[312]

---

303. *Id.*

304. *Id.* at 8.

305. Weedon, et al., *supra* note 298, at 8.

306. *Id.*

307. *Id.* at 4.

308. *Id.* at 8.

309. *Id.* at 10.

310. *Id.*

311. Zameena Mejia, Snapchat Wants to Make Fake News on Its Platform Disappear, Too, *available at* https://qz.com/892774/snapchat-quietly-updates-its-guidelines-to-prevent-fake-news-on-its-discover-platform (last accessed July 25, 2019).

312. *Id.*

Another way to minimize fake news is to strictly enforce the existing real-name policy, which means that users are required to register under their real names or at least provide genuine individual data.[313] In the future, social media companies might be willing to employ artificial intelligence to browse through content in their platforms more effectively and efficiently. In February 2017, Google announced the creation of *Perspective*, which is an artificial intelligence tool capable of finding abusive comments without human assistance.[314] However, it does not delete harmful content, but it merely reports the harmful content to human editors, who will decide whether the given item should or should not be taken down.[315]

Several other groups and entities, independent of social media companies, are also making efforts to try and limit the plague of fake news. In the Philippines, the National Union of Journalists of the Philippines launched a plug-in for Google Chrome designed to block fake news.[316] Known as *Fakeblok*, this plug-in is used to block articles from fake news sites on a person's Facebook newsfeed.[317] This feature also "lets users submit[ ] sites that they believe share fake news."[318] According to Fakeblok — "If you come across something on your Facebook newsfeed that you feel is fake news, you can report it to Fakeblok. A team of journalists will look into your concern. And if verified, the website will be added to the Fakeblok list of sites[.]"[319]

313. The real-name policy means that the social media user is obliged to use his or her real name, the one written on the given person's identification.

314. Madhumita Murgia, *Google Launches Robo-Tool to Flag Up Hate Speech Online*, FIN. TIMES, Feb. 23, 2017, *available at* https://www.ft.com/content/8786cce8-f91e-11e6-bd4e-68d53499ed71 (last accessed July 25, 2019).

315. *Id.*

316. Gelo Gonzales, PH media groups release Facebook fake news blocker, *available at* https://www.rappler.com/technology/news/172310-fake-news-blocker-facebook-philippines-nujp-cmfr-fakeblok (last accessed July 25, 2019).

317. *Id.*

318. *Id.*

319. *Id.*

Another initiative is the *Hoax Analyzer*, a software developed by a team from Indonesia during the Microsoft Imagine Cup.[320] Under this application, "the user simply has to copy the text in question and paste it on the text box of the [application]."[321] It will "then gather[ ] instances of the text or the idea of the text ... found in other websites."[322] "If more than 50% of the sources are classified as 'fact[,]' then the text is declared as 'fact' by the app." Otherwise, [it is] ... a "hoax."[323]

Other fact-checking establishments include the U.S.-based FactCheck.org, [324] Snopes.com, [325] and TruthOrFiction.com. [326] The *old* media have similarly joined this effort as well, with the French newspaper *Le Monde* as an example.[327] It established a special unit called *Les Décodeurs* which not only fact-checks popular stories, but also offers access to *Décodex* — a special database of more than 600 websites identified as sources of fake news.[328]

---

320. Kyle Chua, 'Hoax Analyzer' wins at Microsoft software development contest, *available at* https://www.rappler.com/technology/news/168520-hoax-analyzer-team-cimol-microsoft-imagine-cup-2017 (last accessed July 25, 2019).

321. *Id.*

322. *Id.*

323. *Id.*

324. FactCheck.org, FactCheck Posts, *available at* https://www.factcheck.org/the-factcheck-wire (last accessed July 25, 2019).

325. Snopes, Fact Check, *available at* https://www.snopes.com/fact-check (last accessed July 25, 2019).

326. TruthOrFiction.com, Fact Checks, *available at* https://www.truthorfiction.com/category/fact-checks (last accessed July 25, 2019).

327. Adrien Sénécat, Decodex, a first step towards mass verification of information, *available at* https://www.lemonde.fr/les-decodeurs/article/2017/02/02/le-decodex-un-premier-pas-vers-la-verification-de-masse-de-l-information_5073130_4355770.html (last accessed July 25, 2019) (translate webpage to English by clicking "Translate this page" button in the Google search results).

328. *Id. See also* Décodex, *available at* http://www.lemonde.fr/verification (last accessed July 25, 2019) (translate web page to English by clicking "Translate this page" button in the Google search results).

Other foreign powers have generated some initiatives to fight fake news. StopFake.org, a project created by Ukranian media workers and academia members to debunk Russian propaganda, is one of them.[329] East Stratcom is another one, which is a special unit of the European External Action Service, which also focuses on debunking Russian propaganda.[330] During its 16 months of existence, it has exposed 2,500 false stories.[331] While remarkable, the number is still a drop in the vast ocean. It is still difficult to say whether fact-checking outlets such as *Décodex* will become a mainstream tool to combat fake news. The possibility that only select few users, who are actually less likely to believe fake news, will subscribe to it.

D. Limitations of Self-Imposed Measures

History tells us that removing harmful content published online is not that easy. In fact, an initial evaluation of the application of the abovementioned code of conduct shows that Facebook removed only 28.3% of illegal content within the set timeframe of 24 hours, with Twitter reaching 19.1% and YouTube with 48.5%.[332] Social media companies have achieved very little in

329. Vijai Maheshwari, Ukraine's fight against fake news goes global, *available at* https://www.politico.eu/article/on-the-fake-news-frontline (last accessed July 25, 2019).

330. Questions and Answers about the East StratCom Task Force, *available at* https://eeas.europa.eu/headquarters/headquarters-homepage/2116/-questions-and-answers-about-the-east-stratcom-task-force_en (last accessed July 25, 2019).

331. Mark Scott & Melissa Eddy, *Europe Combats a New Foe of Political Stability: Fake News*, N.Y. TIMES, Feb. 20, 2017, *available at* https://www.nytimes.com/2017/02/20/world/europe/europe-combats-a-new-foe-of-political-stability-fake-news.html (last accessed July 25, 2019).

332. Věra Jourová & Directorate-General for Justice and Consumers, Code of Conduct on countering illegal hate speech online: First results on implementation (Factsheet Released by the European Commission) at 4, *available at* https://ec.europa.eu/information_society/newsroom/image/document/2016-50/factsheet-code-conduct-8_40573.pdf (last accessed July 25, 2019).

countries such as Germany because only 46% of the identified harmful content was removed.[333]

Social media companies admit that there is room for improvement.[334] These companies have made several experiments involving automated content filtering through the use of artificial intelligence.[335] However, it begs the question: what will happen if the algorithm is doubtful or at worst, discriminatory?[336] How can one be sure that the artificial intelligence can clearly distinguish truth from lies? Can transparency be guaranteed despite the possibility of a badly designed algorithm?[337]

It is very difficult to ascertain whether or not the current and preferred measure of social media companies like Facebook to combat fake news, which involves fact-checking and flagging, is enough to meet the demands of society. While it admittedly has positive effects, the reality is that it might not be

333. Deletion of punishable hate comments in the network is not sufficient (Press Release by the *Bundesministerium der Justiz und für Verbraucherschutz* or Germany's Federal Ministry of Justice and Consumer Protection), *available at* https://www.bmjv.de/SharedDocs/Pressemitteilungen/DE/2016/09262016_Hasskriminalitaet.html (last accessed July 25, 2019) (translate web page to English by clicking "Translate this page" button in the Google search results).

334. *See* Chaim Gartenberg, Google News Initiative announced to fight fake news and support journalism, *available at* https://www.theverge.com/2018/3/20/17142788/google-news-initiative-fake-news-journalist-subscriptions (last accessed July 25, 2019) & Tessa Lyons, Increasing Our Efforts to Fight False News, *available at* https://newsroom.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news (last accessed July 25, 2019).

335. *See, e.g.*, Murgia, *supra* note 314 & Daniel Terdiman, Here's How Facebook Uses AI To Detect Many Kinds Of Bad Content, *available at* https://www.fastcompany.com/40566786/heres-how-facebook-uses-ai-to-detect-many-kinds-of-bad-content (last accessed July 25, 2019).

336. *See generally* Ben Wagner, Efficiency vs. Accountability? Algorithms, Big Data and Public Administration, Bureau de Helling, *available at* https://bureaudehelling.nl/artikel-tijdschrift/efficiency-vs-accountability (last accessed July 25, 2019).

337. *Id.*

enough to prevent the proliferation of fake news.[338] Fact-checking becomes an illusory tool for those who patronize fake news in order to further their beliefs and biases, and who evidently distrusts traditional media.[339] More often than not, facts are often swept aside should they contradict one's established opinion. Regardless if it is flagged or not, users will most likely still read it.[340] It is also crucial to note that the existing mechanism on Facebook is open to abuse especially by those who really want to proliferate fake news. These ill-motivated users can simply use the established measure and report genuine information and verified stories as fake. If done in concert with a large number of users, it can potentially destabilize the system and create a crisis within the social media company.

An example of the distrust to these voluntary self-regulating measures is the case involving Cambridge Analytica. In March 2018, the Cambridge Analytica scandal came out.[341] This involves Facebook having exposed data on 50 million users to a researcher who worked at Cambridge Analytica, who at that time, was working for the campaign of Donald Trump.[342] The firm got hold of the data through researcher Aleksandr Kogan, a Russian American who worked at the University of Cambridge.[343] Kogan managed to acquire 50 million user information by creating a Facebook quiz app where it collected

338. Shan Wang, "A Threat to Society": Why a German investigative nonprofit signed on to help monitor hoaxes on Facebook, *available at* http://www.niemanlab.org/2017/02/a-threat-to-society-why-a-german-investigative-nonprofit-signed-on-to-help-monitor-hoaxes-on-facebook (last accessed July 25, 2019).

339. *Id.*

340. *See* Laura Hazard Owen, How to cover pols who lie, and why facts don't always change minds: Updates from the fake-news world, *available at* http://www.niemanlab.org/2017/02/how-to-cover-pols-who-lie-and-why-facts-dont-always-change-minds-updates-from-the-fake-news-world/comment-page-1 (last accessed July 25, 2019).

341. Hanna Kozlowska, et al., The Cambridge Analytica scandal is wildly confusing. This timeline will help, *available at* https://qz.com/1240039/the-cambridge-analytica-scandal-is-confusing-this-timeline-will-help (last accessed July 25, 2019).

342. *Id.*

343. *Id.*

data from those who took it, but not only that, it also allowed the app to collect data from the friends of the quiz takers as well.[344] While Facebook prohibited the sale of this data, Cambridge Analytica proceeded anyway.[345]

The crux of the scandal is less on Cambridge Analytica, but more on Facebook. The question remains as to how much users can trust Facebook with their data. If Facebook cannot even get their act together insofar as data collection is concerned, how much more on the proliferation of fake news on their platform?

V. LAW AND JURISPRUDENCE ON FREEDOM OF EXPRESSION

In resolving the question: "Is fake news, which appears on social media platforms, protected by freedom of expression?," it is important to evaluate how Philippine case law determines if speech is protected or unprotected under the Constitution.

Article III, Section 4 of the 1987 Philippine Constitution is very clear — "No law shall be passed abridging the freedom of speech, of expression, or of the press, or the right of the people to assemble and petition the government for redress of grievances."[346]

In *Chavez v. Gonzales*,[347] the Supreme Court discussed what the freedom principally entails, to wit —

> At the very least, free speech and free press may be identified with the liberty to discuss publicly and truthfully any matter of public interest without censorship and punishment. There is to be no previous restraint on the communication of views or subsequent liability whether in libel suits, prosecution for sedition, or action for damages, or contempt proceedings *unless there be a clear and present danger of substantive evil that Congress has a right to prevent*.[348]

344. *Id.*

345. *Id.*

346. PHIL. CONST. art. III, § 4.

347. Chavez v. Gonzales, 545 SCRA 441 (2008).

348. *Id.* at 483 (citing Gonzales v. Commission on Elections, 27 SCRA 835, 856-57 (1969)) (emphasis supplied).

The Court expounded on why the freedom of expression is "a vital need of a constitutional democracy" and why it was required to ensure stability.[349] Citing *New York Times Co. v. Sullivan*,[350] it declared that the trend is for the State to allow a broad and wide latitude for the exercise of this constitutional freedom.[351]

Chavez v. Gonzales emphasized this by describing that the scope of the freedom is "so broad that it extends protection to nearly all forms of communication."[352] Based on this pronouncement, it can be implied that freedom of expression necessarily extends even to speech performed on social media.

The decision of the court in *Eastern Broadcasting Corporation (DYRE) v. Dans, Jr.*[353] ruled with finality this issue by declaring that such freedom extends to "[a]ll forms of media, whether print or broadcast[.]"[354] Therefore, social media is included despite being a fairly new form of communication.

However, the Court in *Chavez*, made an important distinction that while all forms of media enjoy this freedom, such freedom is not the same on all fours.[355] In fact, the freedom may be lesser in scope compared to others.[356]

The freedom is far from absolute;[357] in fact, Philippine laws allow certain speech to be punished if the freedom is abused. Hence, not all speech is treated alike and some speech may be worse than others which shall not entitle it to protection. For example, under the Philippine jurisdiction, speech constituting

349. *Chavez*, 545 SCRA at 484 (citing *Gonzales*, 27 SCRA at 857).

350. New York Times Co. v. Sullivan, 376 U.S. 254 (1964).

351. *Chavez*, 545 SCRA at 484 (citing *Gonzales*, 27 SCRA at 857).

352. *Chavez*, 545 SCRA at 485.

353. Eastern Broadcasting Corporation (DYRE) v. Dans, Jr., 137 SCRA 628 (1985).

354. *Id.* at 634.

355. *Chavez*, 545 SCRA at 486.

356. *Id.*

357. *Id.*

slander, libel, lewdness and obscenity, and fighting words are unprotected; thus, these forms of speech are subject to punishment.[358]

In determining whether or not speech is protected, Philippine courts have adopted the use of three well-known and practice constitutional tests, namely: (1) dangerous tendency test; (2) balancing of interests test; and (3) clear and present danger test.[359] In a long line of cases, the Philippine Supreme Court has adhered to the clear and present danger test in evaluating whether a speech is entitled to constitutional protection.[360] Under this test, a speech may only be restrained if there it poses substantial danger, and that the speech will likely

---

358. *Id.* (citing 1 HECTOR S. DE LEON, PHILIPPINE CONSTITUTIONAL LAW: PRINCIPLES AND CASES 485 (2003)). It was said —

> Laws have also limited the freedom of speech and of the press, or otherwise affected the media and freedom of expression. The Constitution itself imposes certain limits (such as Article IX on the Commission on Elections, and Article XVI prohibiting foreign media ownership); as do the Revised Penal Code (with provisions on national security, libel and obscenity), the Civil Code (which contains two articles on privacy), the Rules of Court (on the fair administration of justice and contempt) and certain presidential decrees. There is also a 'shield law,' or Republic Act No. 53, as amended by Republic Act No. 1477. Section 1 of this law provides protection for non-disclosure of sources of information, without prejudice to one's liability under civil and criminal laws. The publisher, editor, columnist or duly accredited reporter of a newspaper, magazine or periodical of general circulation cannot be compelled to reveal the source of any information or news report appearing in said publication, if the information was released in confidence to such publisher, editor or reporter unless the court or a Committee of Congress finds that such revelation is demanded by the security of the state.

Id.

359. *Chavez*, 545 SCRA at 487-88.

360. *ABS-CBN Broadcasting Corp.*, 323 SCRA at 825 (citing *Primicias*, 80 Phil.; American Bible Society v. City of Manila, 101 Phil. 386 (1957); Vera v. Arca, 28 SCRA 351 (1969); Navarro v. Villegas, 31 SCRA 730 (1970); Imbong v. Ferrer, 35 SCRA 28 (1970); *Bio Umpar Adiong*, 207 SCRA; & Iglesia Ni Cristo v. Court of Appeals, 259 SCRA 529 (1996)).

lead to an evil that the government "has a right to prevent."[361] It requires that the evil consequences sought to be prevented is substantive, "extremely serious, and the degree of imminence extremely high."[362]

If Congress decides to enact legislation that will punish individuals who publish fake news on social media, the test which should apply is the clear and present danger rule. In the absence of any substantial danger that would lead to an evil sought to be prevented by the State, the content should be allowed and must enjoy constitutional protection.

Further, any law passed by Congress which seeks to punish those who engage in the publication of fake news must be seen with the strictest scrutiny.[363] The government has the burden of overcoming the presumption of unconstitutionality.[364] A law which prohibits a person to speak about something, regardless of the forum, is considered a content-based restriction.[365] This means that the restriction is based on the subject matter of the utterance or speech,[366] which in this case, is fake news. The government has the burden of showing the type of harm the assailed speech would bring especially its gravity and the imminence of the threatened harm.[367] The case

---

361. Cabansag v. Fernandez, et al., 102 Phil. 152, 163 (citing Schenck v. U.S., 249 U.S. 47 (1919)).

362. *Cabansag*, 102 Phil. at 161.

363. *See Chavez*, 545 SCRA at 494 & 496.

364. *Id.* at 494.

365. *Id.*

366. *Chavez*, 545 SCRA at 493.

> Determining if a restriction is content-based is not always obvious. A regulation may be content-neutral on its face but partakes of a content-based restriction in its application, as when it can be shown that the government only enforces the restraint as to prohibit one type of content or viewpoint. In this case, the restriction will be treated as a content-based regulation. The most important part of the time, place, or manner standard is the requirement that the regulation be content-neutral both as written and applied.

Id. at 493 n. 61.

367. *Id.* at 495.

of *Schenck v. United States*[368] is helpful as it tells us that the test is "whether the [speech is] used in such circumstances and are of such [ ] nature as to create a clear and present danger that it will bring about the substantive evils that Congress has a right to prevent."[369]

The Philippine international law obligations call for a similar approach in treating unprotected speech. Article 19 of the UDHR guarantees the right to freedom of expression in broad terms, which include the right "to hold opinions without interference and to seek, receive[,] and impart information and ideas through any media and regardless of frontiers."[370]

The ICCPR also gives legal effect to the rights embodied in the UDHR, which states that everyone shall have the right to freedom of opinion.[371]

Although freedom of expression is a fundamental right, it is not absolute. Article 19 (3) of the ICCPR permits the right to be restricted in the following respects, to wit —

> The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
>
> (a) For respect of the rights or reputations of others;
>
> (b) For the protection of national security or of public order ... , or of public health or morals.[372]

Therefore, content which falls under these two restrictions are no longer protected; hence, they can be subject to regulation and punishment as long as it is provided by law, it is for a legitimate and lawful purpose, and it passes the strict scrutiny test as discussed above.

Therefore, any law passed by Congress which punishes one who speaks fake news must be subjected to the strictest scrutiny. The clear and present danger test is the appropriate test to use considering that the law calls for

368. Schenck v. United States, 249 U.S. 47 (1919).

369. *Id.* at 52.

370. UDHR, *supra* note 33, art. 19.

371. *Id.*

372. ICCPR, *supra* note 40, art. 19, ¶ 3 (a-b).

content-based restriction.[373] No distinction must be made regardless of the medium employed by the speaker or author pursuant to the ruling in *Chavez*. While one can argue that *Chavez* was promulgated in a time where social media was not as popular and prevalent, the Supreme Court acknowledged that the internet, which necessarily includes social networking sites, share similarities with broadcast media;[374] hence, the same standards involving the right to freedom of expression must equally be applied.

But does fake news pose a clear and present danger? There is no jurisprudence discussing fake news and its effects to society. Advocates of free speech argue that despite its false character, such content is fully protected by the Constitution. Yet, the danger that it poses cannot be easily ignored. In the midst of historical revisionism, [375] online troll armies, [376] and political propaganda, [377] the threat to the country's democracy brought by the

373. *See Chavez*, 545 SCRA at 494.

374. *Chavez*, 545 SCRA at 507 (citing Stephen J. Shapiro, *One and the Same: How Internet Non-Regulation Undermines the Rationales Used To Support Broadcast Regulation*, MEDIA L. & POL'Y, Volume VIII, Issue No. 1, at 13).

375. *See* Tara Yap, *Youth urged to resist historical revisionism*, MANILA BULL., Sep. 21, 2018, *available at* https://news.mb.com.ph/2018/09/21/youth-urged-to-resist-historical-revisionism (last accessed July 25, 2019); CNN Philippines, #SuperficialGazette? Netizens slam Official Gazette for 'historical revisionism', *available at* http://nine.cnnphilippines.com/news/2016/09/12/netizens-official-gazette-historical-revisionism.html (last accessed July 25, 2019); & Oscar Franklin Tan, *Why don't we ban historical revisionism?*, PHIL. DAILY INQ., Sep. 19, 2016, *available at* https://opinion.inquirer.net/97450/dont-ban-historical-revisionism (last accessed July 25, 2019).

376. *See* Mikas Matsuzawa, *Duterte camp spent $200,000 for troll army, Oxford study finds*, PHIL. STAR, July 24, 2017, *available at* https://www.philstar.com/headlines/2017/07/24/1721044/duterte-camp-spent-200000-troll-army-oxford-study-finds (last accessed July 25, 2019) & Jonathan Corpus Ong, Trolls for Sale in the World's Social Media Capital, *available at* https://www.asiaglobalonline.hku.hk/philippines-internet-trolls-social-media-duterte (last accessed July 25, 2019).

377. Carolyn Bonquin, Social media influence on Philippines' internet-driven elections, *available at* https://www.cnnphilippines.com/news/2019/4/17/social-media-philippine-elections.html (last accessed July 25, 2019) & Malou Guanzon

proliferation of fake news on social media is real. A discussion of the dangers of fake news is to be made in the following Subsection.

A. The Danger of Fake News in Philippine Society

In analyzing the danger posed by fake news, it is necessary that a discussion of how fake news actually affects society must be made.

The results of the 2016 U.S. Presidential Elections was heavily affected by the proliferation of fake news.[378] According to several reports, fake news regarding the elections were performing better than those which came from genuine news outlets, specifically these fake news sites garnered 8,711,000 hits on Facebook while the latter scored only 7,367,000.[379] A significant fake news item, which was shared over a million times, was the story referring to the alleged endorsement of Pope Francis to Donald Trump[380] — an alarming propaganda story, considering the huge number of Catholics living in the U.S.[381] Most alarming, however, was the fact that it was Donald Trump himself who was posting fake news on his Twitter account such as when he claimed that Barack Obama's birth certificate was fake.[382]

This fake news problem is not unique to the U.S. because even the E.U. has fallen victim to the growing culture of disinformation. During the Brexit

---

Apalisok, *Social media and political propaganda*, CEBU DAILY NEWS, Feb. 24, 2016, *available at* https://cebudailynews.inquirer.net/87190/87190 (last accessed July 25, 2019).

378. *See* Silverman, *supra* note 18.

379. Silverman, *supra* note 18.

380. Tufekci, *supra* note 110.

381. *See* Catholic Review, Percentage of Catholics down but church still largest US denomination, *available at* https://www.archbalt.org/percentage-of-catholics-down-but-church-still-largest-us-denomination (last accessed July 25, 2019).

382. Sapna Maheshwari, *10 Times Trump Spread Fake News*, N.Y. TIMES, Jan. 18, 2017, *available at* https://www.nytimes.com/interactive/2017/business/media/trump-fake-news.html (last accessed July 25, 2019).

referendum campaign,[383] a fake news story saying that taxpayers of Great Britain pay 350 million pounds a week to the EU made its rounds and heavily swung the vote in favor of Brexit.[384] It was so effective that the idea, despite being obviously false, has already influenced the minds of the people and has resulted to the Brexit.[385]

In the Philippines, the danger brought by fake news has seeped through the very foundation of the Philippine society. The disinformation is deeply rooted in the various propaganda mechanisms set up by different parties. An example of this disinformation campaign is when at least one anonymous Facebook account shared a March 2016 story of Rappler with the title "Man with bomb nabbed at Davao checkpoint."[386] This post was made after the Davao bombing incident.[387] This post was rapidly shared through various Facebook pages in support of President Duterte.[388] It caused serious disinformation considering that it made people believe that the man who was in possession of the bomb was captured the day after the September 2016 bombing incident, when in fact the article was talking about something else which happened in March 2016.[389] This created an altered reality for readers

383. *See generally* Brian Wheeler, et al., Brexit: All you need to know about the UK leaving the EU, *available at* https://www.bbc.com/news/uk-politics-32810887 (last accessed July 25, 2019).

384. Jon Stone, British public still believe Vote Leave '£350million a week to EU' myth from Brexit referendum, *available at* https://www.independent.co.uk/news/uk/politics/vote-leave-brexit-lies-eu-pay-money-remain-poll-boris-johnson-a8603646.html (last accessed July 25, 2019).

385. *See* John Lichfield, *Boris Johnson's £350m claim is devious and bogus. Here's why*, GUARDIAN, Sep. 18, 2017, *available at* https://www.theguardian.com/commentisfree/2017/sep/18/boris-johnson-350-million-claim-bogus-foreign-secretary (last accessed July 25, 2019).

386. Ressa, *supra* note 24. *See also* Editha Caduaya, Man with bomb nabbed at Davao checkpoint, *available at* https://www.rappler.com/nation/127132-man-bomb-nabbed-davao-checkpoint (last accessed July 25, 2019).

387. Ressa, *supra* note 24.

388. *Id.*

389. *Id.*

on Facebook, which made people believe that the actions made by President Duterte was justified after the declaration of a state of lawlessness.[390] This was so effective that the story stayed on the top ten stories in Rappler's website for more than two days.[391]

Another campaign tactic was employed by Peter Tiu Laviña, President Duterte's then spokesman, who used a picture taken from Brazil to justify the war on drugs of the President.[392]

These incidents started to blow up when they went far worse from these confusion tactics to actual harassment or historical revisionism through the form of fake news. The National Union of Journalists of the Philippines has made calls to the government to investigate social media attacks against journalists.[393]

The problem with fake news is rooted not only in the effects created not only to the individual who is the target of the falsity, but also in its long-term effects to both society and democracy. Historical revisionism has also become prevalent by reason of fake news. An example of this is when the Official Gazette posted a photo of former President Ferdinand Marcos on Facebook where a part of the caption reads — "In 1986, Marcos stepped down from the presidency to avoid bloodshed during the uprising that came to be known as 'People Power.'"[394]

---

390. *See* Ressa, *supra* note 24.

391. *Id.*

392. Camille Elemia, FACT CHECK: Photo used by Duterte camp to hit critics taken in Brazil, not PH, *available at* https://www.rappler.com/nation/144551-duterte-camp-brazil-photo-rape-victim-critics (last accessed July 25, 2019).

393. Rappler.com, NUJP to Palace: Investigate social media attacks vs journalists, *available at* https://www.rappler.com/nation/146674-nujp-andanar-investigate-social-media-attacks-journalists (last accessed July 25, 2019).

394. Marlon Ramos & Yuji Vincent Gonzales, *Gazette draws flak for Marcos boo-boo*, PHIL. DAILY INQ., Sep. 13, 2016, *available at* https://newsinfo.inquirer.net/814843/gazette-draws-flak-for-marcos-boo-boo (last accessed July 25, 2019) & Rappler.com, Official Gazette under fire for Marcos photo caption, *available at* https://www.rappler.com/nation/145920-philippines-official-gazette-ferdinand-marcos-photo-historical-revisionism (last accessed July 25, 2019).

This post received a lot of backlash because of how the post seemed to rewrite what history dictates.[395] Commenters were arguing that the post was discounting the suffering of Filipinos during the martial law era of Marcos.[396]

Another staunch defender of President Duterte is Mocha Uson, who is infamous for her personal blog, which many describe as a source of fake news in the country.[397] Uson, who was then a public official,[398] made a poll on her Facebook page saying, "Naniniwala ba kayo na ang 1986 EDSA PEOPLE POWER ay isang produkto ng FAKE NEWS???" (Do you believe that the 1986 EDSA People Power is a product of fake news?)[399] Uson's poll resulted to 84% agreeing to her question.[400] While this result is unfortunate and depressing as a nation, this act performed by Uson is questionable and downright unethical for a public officer. The fact that the EDSA People Power happened to overthrow a dictator is a historical fact not up for debate.[401] Not

395. Ramos & Gonzales, *supra* note 394; Rappler.com, *supra* note 393; & CNN Philippines, *supra* note 375.

396. Rappler.com, *supra* note 393.

397. *See* Vera Files, VERA FILES FACT SHEET: A trail of false claims made and fake news shared by Mocha Uson, *available at* https://verafiles.org/articles/vera-files-fact-sheet-trail-false-claims-made-and-fake-news (last accessed July 25, 2019) & Don Kevin Hapal & Bonz Magsambol, MOCHA USON: FAKE NEWS VICTIM OR FAKE NEWS PEDDLER?, *available at* https://www.rappler.com/newsbreak/investigative/185560-mocha-uson-posts-news (last accessed July 25, 2019).

398. *Id*.

399. MOCHA USON BLOG, Poll, *Naniniwala ba kayo na ang 1986 EDSA PEOPLE POWER ay isang produkto ng FAKE NEWS???*, Feb. 25, 2018: 2:24 a.m., FACEBOOK, *available at* https://www.facebook.com/Mochablogger/posts/naniniwala-ba-kayo-na-ang-1986-edsa-people-power-ay-isang-produkto-ng-fake-news/10156335785651522 (last accessed July 25, 2019).

400. *Id*.

401. Miguel Escobar, *The EDSA Revolution Happened, No Matter What Your Poll Says*, ESQUIRE, Feb. 26, 2018, *available at* https://www.esquiremag.ph/politics/fck-your-internet-poll-people-power-happened-a00207-20180226 (last accessed July 25, 2019).

only is this written in Philippine history books,[402] this is also etched in the memories of the Filipino people. This *poll* made by Uson constitutes, in the opinion of the Authors, a form of historical revision, which is brought about by the culture of fake news and disinformation in the Philippines.

The danger it poses is quite clear — what will happen if people start to believe that the EDSA People Power was indeed a product of fake news? Let alone, this declaration was made by no less than a public official. It is quite obvious that content like this, although not libelous or slanderous, is a matter so damning to national interest, specifically to Philippine history, that it does not deserve the protection of the Constitution.

VI. THE LIABILITY OF ONLINE INTERMEDIARIES

With over seven billion people connected to the Internet, [403] online intermediaries play a vital role in the fight against fake news. An online intermediary connotes a broad meaning. It may refer to web hosting companies, internet service providers (ISPs), and social media companies.[404]

History would tell us that online intermediaries "were generally subject to limited regulation[.]"[405] In recent years, more pressure has been put on online intermediaries to ensure that their platforms are not used as avenues to publish and disseminate unprotected speech such as defamation, slander, or obscenity. [406] In fact, more governments have initiated ways to either encourage or even compel these intermediaries to remove or block content which they believe is harmful or unprotected.[407] In most cases, States have seemingly obligated these online intermediaries to serve as the government's police to block, modify, or delete harmful content on their platforms.[408] This

402. *Id.*

403. ARTICLE 19, INTERNET INTERMEDIARIES: DILEMMA OF LIABILITY 3 (2013).

404. *Id.* (citing Organisation for Economic Cooperation and Development, The Economic and Social Role of internet Intermediaries at 9, *available at* https://www.oecd.org/sti/ieconomy/44949023.pdf (last accessed July 25, 2019)).

405. ARTICLE 19, *supra* note 403, at 3.

406. *See* ARTICLE 19, *supra* note 403, at 3.

407. ARTICLE 19, *supra* note 403, at 3.

408. *Id.*

pressure is not limited to social media companies. Companies such as eBay or Paypal can also be subject to this kind of pressure[409] — all of which may eventually lead to government censorship.

While it is true that online intermediaries such as social media companies have voluntary measures to police harmful content on their platforms, the lack of transparency and clear standards is quite dangerous. This shows that when users engage in an online contract to use the social media platform, there is an increased risk of regulation and censorship which is subject to limited transparency and accountability.

While it is important to determine whether intermediaries are liable, it is equally vital to understand what the different types of intermediaries are. According to a study conducted by Article 19, there are three distinct models of liability for online intermediaries: strict liability, safe harbor, and broad immunity.[410]

Under the *strict liability* model, "internet intermediaries are liable for third-party content."[411] Countries such as Thailand and China employ this model.[412] Intermediaries are required by law to police and control content on their platforms; otherwise, they will face sanctions, including the revocation of business license and/or the imposition of criminal penalties.[413]

The *safe harbor* model is another interesting model wherein intermediaries are granted immunity, "*provided* they [conform to] certain requirements."[414] This model is said to be "at the heart of the [ ]*notice and take down*[ ] procedure[.]"[415] There are two approaches under this model: the vertical and horizontal approach.[416] Under the vertical approach, "[t]he liability regime

409. *See* ARTICLE 19, *supra* note 403, at 3.

410. ARTICLE 19, *supra* note 403, at 7.

411. *Id.*

412. *Id.*

413. *Id.*

414. *Id.* (emphasis supplied).

415. *Id.* (emphasis supplied).

416. ARTICLE 19, *supra* note 403, at 7.

only applies to certain types of content"[417] such as in cases of copyright infringement or obscenity.[418] In the horizontal approach, on the other hand, immunity is granted on various levels.[419] If the intermediary serves only to provide technical access to the internet, such as a conduit, complete immunity attaches.[420] However, if the intermediary "fail[s] to act 'expeditiously' to remove or disable access to [ ]illegal[ ] information"[421] when they know about it, the immunity is lost.[422] This is exactly the spirit of the *notice and take down* procedure.

Lastly, the *broad immunity* model grants either broad or conditional immunity from liability for third-party content and exempts them from the responsibility to monitor content in its platform.[423] Under this model, intermediaries are merely "messengers," and therefore, are not responsible for the content in their platform.[424] They are not considered "publishers" for the content distributed through the intermediary.[425] This is the model followed by the U.S. and the E.U.[426]

With these three models in mind, the question is: "Should online intermediaries be held liable for content in their platforms?"

In 2011, the Joint Declaration on Freedom of Expression and the Internet, recommended that

---

417. *Id.*

418. *Id.* (citing Digital Millennium Copyright Act, H.R. No. 2281, 105th Cong., 2d Sess. (1997-1998) (U.S.)).

419. ARTICLE 19, *supra* note 403, at 7.

420. *Id.* (citing E.U. E-Commerce Directive, *supra* note 234, at 12-13).

421. ARTICLE 19, *supra* note 403, at 7 (citing E.U. E-Commerce Directive, *supra* note 234, at 13).

422. *Id.*

423. ARTICLE 19, *supra* note 403, at 7

424. *Id.*

425. *Id.*

426. *Id.*

> No one should be liable for content produced by others when providing technical services, such as providing access, searching for, or transmission or caching of information;
>
> Liability should only be incurred if the intermediary has specifically intervened in the content, which is published online;
>
> ISPs and other intermediaries should only be required to take down content following a court order, contrary to the practice of notice and [take down].[427]

In 2011, the United Nations (UN) Special Rapporteur on freedom of expression also said: "Censorship measures should never be delegated to a private entity, and [ ] no one should be held liable for content on the Internet of which they are not the author. Indeed, no State should use or force intermediaries to undertake censorship on its behalf[.]"[428]

Further, the said UN Special Rapporteur recommended that "intermediaries should only implement restrictions to these rights after judicial intervention."[429] In fact, international bodies have criticized the *notice and take down* procedure.[430] The 2011 OSCE report on Freedom of Expression on the Internet stated —

> Liability provisions for service providers are not always clear and complex notice and take[ ]down provisions exist for content removal from the Internet within a number of participating States. Approximately, 30 participating States have laws based on the [E.U.] E-Commerce Directive.

---

427. ARTICLE 19, *supra* note 403, at 10-11 (citing Joint Declaration on Freedom of Expression and the Internet, *available at* https://www.article19.org/data/files/pdfs/press/international-mechanisms-for-promoting-freedom-of-expression.pdf (last accessed July 25, 2019) [hereinafter Joint Declaration]).

428. ARTICLE 19, *supra* note 403, at 11 (citing Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression*, ¶ 43, 17th Session of the Human Rights Council, U.N. Doc. A/HRC/17/27 (May 16, 2011) (by Frank La Rue)).

429. ARTICLE 19, *supra* note 403, at 11. (citing Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *supra* note 428, ¶ 47).

430. ARTICLE 19, *supra* note 403, at 11.

> However, the [E.U.] Directive provisions rather than aligning state-level policies, created differences in interpretation during the national implementation process. These differences emerged once the provisions were applied by the national courts.[431]

The danger of the *notice and take down* procedure without any form of judicial intervention lies with the reality that intermediaries may choose to always take down content which they have been notified to be harmful, when in reality, it may actually be protected speech.[432] Under the *notice and take down* procedure, the determination of what is lawful or unlawful is placed on the shoulders of the intermediaries — a situation that is very difficult and nuanced because their primary interest is not to the people, but to themselves.[433]

Despite this, many States believe that intermediaries should become liable to a certain extent for harmful content on their platforms on two reasons: First, it is practical to hold intermediaries liable because they are at the best position to actually police the harmful content;[434] and second, it is fair to expect these intermediaries to ensure that their platform is not used as an avenue to spread or disseminate fake news and other harmful content, considering that they actually benefit commercially by reason of the engagement brought about by their users.[435]

This Section will now proceed to discuss international jurisprudence concerning the liability of online intermediaries.

In *Hermès International v. eBay*,[436] the French Civil Court of Troyes held eBay liable after it found its efforts to suppress counterfeit items insufficient,

---

431. *Id.* (citing YAMAN AKNEDIZ, FREEDOM OF EXPRESSION ON THE INTERNET 47 (2012)).

432. *See* ARTICLE 19, *supra* note 403, at 11.

433. *Id.* (citing Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, *supra* note 428, ¶ 42).

434. ARTICLE 19, *supra* note 403, at 14.

435. *Id.*

436. Hermès International v. eBay, No. 06/02604 (Tribunaux de grande instance [TGI] [ordinary court of original jurisdiction]) (2008) (unreported) (Fr.).

which in this case, was a luxury bag.[437] eBay was made to pay €20,000 in damages.[438] While this case involved infringement of intellectual property, it is illustrative insofar how European courts assess the liability of online intermediaries.

Delfi AS v. Estonia[439] is another interesting case because it was the first time that the European Court of Human Rights ruled on the issue of intermediary liability.[440] The court held Delfi liable for the comments made by users in its platform.[441]

"Delfi is one of the largest news portals ... in Estonia" and it "publishe[s] up to 330 news articles a day[.]"[442] Its comments section is infamous for having offensive comments published by users, who may even post anonymously.[443] Delfi does not edit these comments.[444]

In January 2006, Delfi published an article with a headline, "SLK Destroyed Planned Ice Road." [445] It generated 185 comments, 20 of which "contained personal threats and offensive language" against a member of SLK's supervisory board.[446]

---

437. Todd Evan Lerner, *Playing the Blame Game, Online: Who is Liable when Counterfeit Goods are Sold Through Online Auction Houses?*, 22 PACE INT'L L. REV. 241, 248-49 (2010).

438. *Id.* at 243.

439. Delfi AS v. Estonia, Application No. 64569/09, Judgment (Eur. Ct. H.R. June 16, 2015).

440. *Id.* ¶ 111.

441. *Id.* ¶¶ 156 & 162.

442. *Id.* ¶ 11.

443. *Id.* ¶¶ 12 & 15.

444. *See Delfi AS*, Application No. 64569/09, ¶ 154.

445. *Delfi AS*, Application No. 64569/09, ¶ 16.

446. *Id.* ¶ 17.

This member eventually requested Delfi to take down the 20 offensive comments and to pay him damages.[447] Delfi removed the comments but ignored the demand for damages.[448]

The European Court of Human Rights affirmed the decision of the Estonian Supreme Court as regards Delfi's liability because: (1) the comments were tantamount to hate speech which was not protected under international law;[449] and (2) although the users may be held liable under domestic law, [450] problems as to their identification becomes problematic, and given Delfi's superiority over their users, it should be made liable.[451]

The court emphasized that although the general rule is that no person should be made to answer for acts which he did not commit, Delfi remains to be liable under the *economic interest test*, and the *control test*, to wit —

> [I]n the comments section, the applicant company actively called for comments on the news items appearing on the portal. The number of visits to the applicant company's portal depended on the number of comments; the revenue earned from advertisements published on the portal, in turn, depended on the number of visits. Thus, the [Estonian] Supreme Court concluded that the applicant company had an economic interest in the posting of comments. In the view of the [Estonian] Supreme Court, the fact that the applicant company was not the writer of the comments did not mean that it had no control over the comments section.[452]

It is important to note that Delfi was not punished because of content which it had editorial control over, but because of content posted by users in the comments section. Therefore, if the test is not actually whether or not the intermediary has control over the content, but rather if it has commercial interest and control over it, the risk of encouraging intermediaries to act as censor over content published in their platforms increases — which in the

447. *Id.* ¶ 18.

448. *Id.* ¶¶ 19-20.

449. *Id.* ¶¶ 110, 114, & 115.

450. *Id.* ¶ 128.

451. *See Delfi AS*, Application No. 64569/09, ¶¶ 129; 144-45; & 161

452.*Delfi AS*, Application No. 64569/09, ¶ 144.

long run, may even amount to prior restraint — clearly becoming a violation of the right to freedom of expression.

Very recently, the Supreme Court of Canada issued a global injunction against Google,[453] which is a passive intermediary, to de-list a particular website worldwide.[454]

In these cases, one can observe that intermediary liability can arise in two instances: (1) if the intermediary exercises editorial control over the content; and (2) if the intermediary derives economic benefit over the content and despite its harmfulness, the intermediary decides to retain it.

With these in mind, the following Sections will discuss whether online intermediaries should be held liable considering the present social realities in the Philippines.

VII. COMPARATIVE ANALYSIS: THE VARIOUS STATE APPROACHES IN ADDRESSING THE FAKE NEWS PROBLEM

The Philippines is not alone in searching for solutions to fight the proliferation of fake news. In this Section, the Authors will discuss the methods and means employed by various countries to see how the Philippines can learn from them and if there is any effective mechanism that can be emulated in the society. Germany, one of the pioneer countries in enacting anti-fake news legislation, is the primary model of this Section. In addition, countries including Russia, Singapore, Venezuela, Kenya, and the United Kingdom will also be evaluated.

At the end of June 2017, Germany passed the *Netzwerkdurchsetzungsgesetz* law, otherwise known as the Network Enforcement Act or the NetzDG.[455] This law aims to punish social media companies that do not remove *illegal* content within 24 hours, which can be extended up to seven days for highly complex cases.[456] These companies can face fines up to 50 million euros should the social media company choose not to remove such harmful

453. Google Inc. v Equustek Solutions Inc., 2017 SCC 34, ¶ 17 (2017) (Can.)

454. *See Google Inc.*, 2017 SCC, ¶ 16.

455. BBC News, Germany starts enforcing hate speech law, *available at* http://www.bbc.com/news/technology-42510868 (last accessed July 25, 2019).

456. *Id.*

content.[457] The law further requires a social media company to establish a *comprehensive complaints structure* to allow quick reporting to its team of the assailed content.[458] According to reports, Facebook recruited additional staff to deal with reports of violations of the NetzDG.[459]

Deemed one of the most controversial and strictest laws of free speech yet,[460] the NetzDG is a firm example as to how far a government can go to fight back fake news. *Bild*, Germany's biggest newspaper, has called for the abolition of the NetzDG stating that the law against online hate speech "has already failed on its very first day. It [should be] abolished immediately."[461] This was quickly rebutted by Germany's justice minister, Heiko Mass, saying, "Incitement to murder, threats, [and] insults[,] and incitement of the masses or Auschwitz lies are not an expression of freedom of opinion but rather attacks on the freedom of opinion of others."[462] This exchange arose from the temporary suspension of Beatrix von Storch, Deputy Leader of Alternative für Deutschland (AfD), from Twitter earlier in 2018.[463]

457. *Id.*

458. *Id.*

459. *Id.*

460. *See* Lucinda Southern, Critics say Germany's hate speech law comes at a price, *available* https://digiday.com/media/like-taking-sledgehammer-fix-wristwatch-critics-say-germanys-hate-speech-law-comes-price (last accessed July 25, 2019).

461. Julian Reichelt, *Giant dispute over blocked Twitter accounts and Facebook deletions*, BILD, Mar. 1, 2018, *available at* https://www.bild.de/politik/inland/gesetze/kommt-jetzt-die-meinungspolizei-54367844.bild.html (last accessed July 25, 2019) (translate web page to English by clicking "Translate this page" button in the Google search results).

462. Philip Oltermann, *Tough new German law puts tech firms and free speech in spotlight*, GUARDIAN, Jan. 5, 2018, *available at* https://www.theguardian.com/world/2018/jan/05/tough-new-german-law-puts-tech-firms-and-free-speech-in-spotlight (last accessed July 25, 2019).

463. *Id.*

The Human Rights Watch commented on the law in a recent report.[464] It opined that while "[c]ompanies must inform users of all decisions made in response to complaints and provide justification,"[465] the law does not provide for "judicial oversight or a [judicial] process ... when users want to [question]" the action performed by the social media company.[466]

Richard Allan, Facebook's Vice President for Europe, the Middle East and Africa (EMEA) Public Policy, gave a strong statement that "[p]eople think deleting illegal content is easy but [it is] not."[467] He further goes on saying that Facebook reviews every NetzDG report meticulously and that they make consultations with their legal experts to ensure full compliance with the law.[468]

Johannes Ferchner, spokesman on justice and consumer protection for the Social Democracts and one of the authors of the law, assured that there will be an inclusion of a proviso in the NetzDG that will provide users the legal opportunity to have unjustified removal of content restored.[469] In fact, this appears to be a sensible solution considering that too much content was being deleted according to Thomas Jarzombek, a Christian Democrat.[470]

With Germany leading the charge, other countries have used the NetzDG as an example and statutory model to enact laws or propose measures to combat fake news in their countries.[471] Singapore, for example, recently

---

464. Human Rights Watch, Germany: Flawed Social Media Law, *available at* https://www.hrw.org/news/2018/02/14/germany-flawed-social-media-law (last accessed July 25, 2019).

465. *Id.*

466. *Id.*

467. Emma Thomasson, Germany looks to revise social media law as Europe watches, *available at* https://www.reuters.com/article/us-germany-hatespeech/germany-looks-to-revise-social-media-law-as-europe-watches-idUSKCN1GK1BN (last accessed July 25, 2019).

468. *Id.*

469. *Id.*

470. *Id.*

471. Human Rights Watch, *supra* note 464.

passed a law that punishes those who spread fake news with Germany as its model.[472]

In Russia, the State Duma also deliberated on two versions of a law as to how to regulate content on social media.[473] A news report stated that the lawmakers made reference to Germany's NetzDG.[474] Under these proposals, social media companies "with more than [two] million registered users and other 'organizers of information dissemination' in Russia" are obligated to remove illegal content, within 24 hours upon receiving a complaint.[475] The term "illegal content" includes those that propagates war, incites national, racial, or religious hatred, defamation, or a violation of other existing laws.[476]

On November 2017, Venezuela enacted the "Anti-Hate Law for Peaceful Coexistence and Tolerance."[477] This law imposes high fines on social media companies that "fail to delete content that 'constitute[s] propaganda advocating war or national, racial, religious, political, or any other kind of hatred,' within six hours of posting."[478]

The Communications Authority of Kenya issued guidelines in July 2017 that requires social media companies to close accounts that are utilized in the dissemination of "undesirable political contents" within 24 hours once the

472. Mansi Jaswal, From Singapore to France: These countries have created laws to fight fake news, *available at* https://www.businesstoday.in/current/world/fake-news-fake-news-law-singapore-fake-news-law-countries-that-have-fake-news-law-/story/345144.html (last accessed July 25, 2019). *See also* Human Rights Watch, *supra* note 464.

473. *United Russia Tries to Fight 'Fake News' (In Its Own Way)*, MOSCOW TIMES, July 13, 2017, *available at* https://www.themoscowtimes.com/2017/07/13/united-russia-tries-to-fight-fake-news-a58376 (last accessed July 25, 2019).

474. *Id.*

475. Human Rights Watch, *supra* note 464.

476. *Id.*

477. *Id.*

478. *Id.*

social media company is notified.[479] However, there is still no record of any person punished under these guidelines as of 2017.[480]

The European Commission has also expressed its firm position to require social media companies to take on greater responsibility for the removal of harmful content on their platforms. This has led to the formulation of a code of conduct for Information Technology (IT) companies.[481] This has also resulted to both the U.K. and French governments working hand in hand in the creation of a joint action plan to increase effectiveness in identifying and deleting harmful content posted online.[482]

Lastly, in the United Kingdom, one of the ministers of Prime Minister Theresa May suggested that tax penalties be imposed against social media companies that were shown to be "slow" in the removal of content published on their platform.[483] The Prime Minister herself has made her position clear

479. *Id*. *See generally* Guidelines on Prevention of Dissemination of Undesirable Bulk and Premium Rate Political Messages and Political Social Media Content via Electronic Communications Networks (A Publication by the National Cohesion and Integration Commission Kenya and the Communications Authority of Kenya), *available at* https://ca.go.ke/wp-content/uploads/2018/02/Guidelines-on-Prevention-of-Dissemination-of-Undesirable-Bulk-and-Premium-Rate-Political-Messages-and-Political-Social-Media-Content-Via-Electronic-Networks-1.pdf (last accessed July 25, 2019).

480. *See* Freedom House, Freedom on the Net 2017 - Kenya, *available at* https://www.refworld.org/docid/5a547d2fa.html (last accessed July 25, 2019).

481. European Commission, Code of Conduct on Countering Illegal Hate Speech Online, *available at* https://ec.europa.eu/newsroom/document.cfm?doc_id=42985 (last accessed July 25, 2019) (click the "en" option to download the English version of the Code of Conduct).

482. French-British Action Plan, *available at* https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/619333/french_british_action_plan_paris_13_june_2017.pdf (last accessed July 25, 2019).

483. BBC News, Call for tech giants to face taxes over extremist content, *available at* http://www.bbc.co.uk/news/uk-42526271 (last accessed July 25, 2019).

that social media companies should do more in the fight against terrorism by removing such content online.[484]

Taking everything in consideration, it can be observed that Germany is clearly an educational model to be used for a State that seeks to impose additional legislation to regulate content on social media. It is further observed that in these cases, except in Singapore, the focus is on the social media company having the responsibility to remove the harmful content on their platform, rather the individual.

What is interesting in Germany's model is the lack of judicial review should the social media company's action be unjustified.[485] One of the creators of the NetzDG has already admitted the possibility of the inclusion of that mechanism in their law.[486] It is worth noting that according to the Human Rights Watch report, one of the main downfalls of the NetzDG is the lack of an avenue of the author of the content to seek judicial intervention in cases of arbitrary deletion or correction made by the social media company.[487] This must be remembered when coming up with a proposal later on in this Article.

VIII. Determining the Best Approach in Addressing the Fake News Phenomenon in the Philippine Society

A. Punishing the Architects of Disinformation

The proposal of punishing the user or author for creating and publishing the fake news on social media faces a great challenge in light of the constitutional right to freedom of expression. Advocates of regulation can argue that it is akin to libel or slander, hence a form of subsequent punishment. However, libel and slander are recognized exceptions to protected speech because law and jurisprudence have recognized that both have passed strict scrutiny and publication satisfies the *clear and present danger* rule.

---

484. Heather Stewart & Jessica Elgot, *May calls on social media giants to do more to tackle terrorism*, GUARDIAN, Jan. 24, 2018, *available at* https://www.theguardian.com/business/2018/jan/24/theresa-may-calls-on-social-media-giants-to-do-more-to-tackle-terrorism (last accessed July 25, 2019).

485. Human Rights Watch, *supra* note 464.

486. Thomasson, *supra* note 467.

487. Human Rights Watch, *supra* note 464.

On the other hand, there is yet to be a Supreme Court case or a domestic law which specifically provides that liability be imposed on the author of the fake news posted on social media. There is also currently no law which defines what is fake news or not. The standards in determining what are fake news not protected by the Constitution remain unclear. While it can be argued that Article 147 of the Revised Penal Code can be applied to social media platforms pursuant to the ruling in *Disini, Jr.*, the strength of this argument is yet to be affirmed by the Court.

To regulate the author or speaker is akin to content-based restriction. Any law imposing such restriction must satisfy the *strict scrutiny* test. It must appear that the gravity of the fake news must pass the *clear and present danger* test; otherwise, it will be declared unconstitutional.

Recently, Senator Grace Poe filed Senate Bill No. 1680 which seeks to punish government officials who publish or disseminate fake news.[488] According to Senator Poe, the bill is important considering that public officials are demanded to be accountable at all times for all their official acts; they should not spearhead the publication and dissemination of fake news in the country.[489]

Professor Florin Hilbay, a respected constitutionalist in the country, supports this "higher standard" according to his opinion during the Senate Investigation on fake news, to wit —

> [F]alse information provided by public officials poses 'special problems:'
>
> (1) They are paid with public funds. It is an outrage that they receive taxpayers' money so they can lie[;]
>
> (2) The official status provides official imprimatur to false information, whether posted in private or official social media accounts[;]
>
> (3) Their public employment provides them with access to government facilities creating a semblance of credibility where otherwise[,] there might be none[;] [and]

488. Michael Bueza, Poe on bill vs 'fake news': Hold gov't officials to higher standard, *available at* https://www.rappler.com/nation/195733-poe-bill-vs-fake-news-roque-govt-officials-higher-standards (last accessed July 25, 2019).

489. *Id.*

(4) Their access to government facilities means that the false information they provide gets widely distributed.[490]

This bill stirred some controversy. Harry Roque, then the spokesperson of the President, argued during the Senate hearing that the bill is presumed to be unconstitutional.[491] He cites both the equal protection clause and the right to freedom of expression as his two main reasons why the bill is unconstitutional.[492]

While the effort of Senator Poe in filing Senate Bill No. 1680 is commendable, Roque's position is more in line with both the Constitution and jurisprudence. First of all, who is to be the standard of truth? In case of libel and slander, the standard appears to be clear. When one is attacked either orally or in writing, with an intent to defame one's reputation, there can be no question that the act committed is libel or slander, as the case may be.[493] However, in the case of fake news, the line is not so clearly drawn. Must there be an intent to actually publish the fake news? What if the content refers to more than one possible truth? Who is to say what is true and what is not? Unless all these questions are addressed, and sufficient standards are put in place, it becomes almost impossible to regulate the speech of an individual. If done, it would be akin to putting a tape on the mouth of the person because the chilling effect will be so great that clearly, it violates the Constitution.

More importantly, the disinformation infrastructure in the Philippines goes far beyond the actual speaker; in reality, it is more complex than it seems. In a recent study entitled *Architects of Networked Disinformation*, it emphasized

490. Interaksyon, As fake news proliferates, former SolGen wants 'information police for govt officials' created, *available at* http://www.interaksyon.com/breaking-news/2017/10/04/101546/as-fake-news-proliferates-former-solgen-wants-information-police-for-govt-officials-created (last accessed July 25, 2019).

491. John Ted Cordero/LBG, GMA News, Poe bill vs. spreading of fake news presumed unconstitutional —Roque, *available at* https://www.gmanetwork.com/news/news/nation/642874/poe-bill-vs-spreading-of-fake-news-presumed-unconstitutional-roque/story (last accessed July 25, 2019).

492. Christine O. Avendaño, *Poe, Roque clash over fake news bill*, PHIL. DAILY INQ., Mar. 16, 2018, *available at* http://newsinfo.inquirer.net/975738/poe-roque-clash-over-fake-news-bill (last accessed July 25, 2019).

493. *See* REVISED PENAL CODE, art. 355.

how disinformation in the Philippines is a product of collective and professionalized production.[494] It pointed out that the chief disinformation architects behind this structured approach in shaping public opinion are usually the public relations analysts who are well-versed with digital media.[495] They work hand in hand with groups of people acting as *trolls* who serve to create an illusion of engagement.[496] Clearly, the famous personalities who are seen to spread and peddle fake news on social media is basically just the tip of the iceberg. Its complexity is brought about by the various parties involved, contrary to popular belief that disinformation is brought about by only one person alone.[497] Therefore, the need to create solutions to address this well-organized system of disinformation is clear as day; otherwise, its effects will continue to pervade the society.

In view of the foregoing, it becomes apparent that imposing any form of liability, whether civil or criminal, to the author or architect of fake news is very difficult considering the lack of any adequate and sufficient standard in determining liability. The absence of any clear definition of what constitutes fake news is the main source of complexity, but even assuming a well-encompassing definition is crafted, the line between the right of a person to speak his or her own opinion and the power of the State to impose liability is very thin, which undeniably creates a strong chilling effect that can effectively discourage people to even speak up — a reality that is present today, and one that is incompatible with the concept of democracy.

B. Online Intermediary Liability and Accountability

State-imposed legislation is at times necessary to protect public interest. It would be an easier situation if social media companies are treated the same way as traditional media companies like print and broadcast media in the

494. Jonathan Corpus Ong & Jason Vincent A. Cabañes, Architects of Networked Information: Behind the Scenes of Troll Accounts and Fake News Production in the Philippines (A Report Produced by Newton Tech4Dev Network) at 2 & 7, *available at* http://newtontechfordev.com/wp-content/uploads/2018/02/Architects-of-Networked-Disinformation-Executive-Summary-Final.pdf (last accessed July 25, 2019).

495. *Id.* at 2 & 5.

496. *Id.* at 6.

497. *See* Ong & Cabañes, *supra* note 494, at 5-7.

Philippines. In the Senate Investigations, Armand Dean Nocum, a public relations practitioner and a resource person, said —

> *Ang sagot ko diyan*, Senator, *nakikita po na tulad noong sinabi ni* Madam Chair *kanina po sa* introduction, *may mga* pertinent laws *na po tayo*. In fact, *naka-*apply *na iyon po sa* old media, *sa* Rappler, *sa lahat*, *sa* ABS-CBN. *Siguro po iyong* application *lang talaga, iyong* libel, Cybercrime Law. *Ang masakit nga ho*, fully applied *sa* old media *pero sa* new media *po parang walang* application of any kind *po kaya* any blogger can say, 'Blogger *na ako*, this is my opinion and I can hit Senator Manny Pacquiao if I want.' *Parang ganoon po*.[498]

However, there may be an argument to the effect that if social media companies be treated similarly, then it must be considered as a mass media company; hence, it should be subject to the constitutional requirement of being wholly Filipino-owned under the Constitution.[499]

This argument does not hold water in light of the nature of social media companies. They are not mass media entities. What distinguishes the two is that in mass media, the audience is in a passive position where he is only in the receiving end of the communication channel, while in social media, the audience is at the center and is both the creator and the audience, creating that social experience characterized by collaboration and interaction.[500] The Constitution does not provide an exact definition of what is mass media; however, the Department of Justice in 1998 defined it as "any medium of communication designed to reach the masses and that tends to set the standards, ideals[,] and aims of the masses, the distinctive feature of which is the dissemination of information and ideas to the public, or a portion

---

498. Rappler, Video, *Part 2: Senate Hearing on Fake News, 30 January 2018*, Jan. 30, 2018, YOUTUBE, *available at* https://www.youtube.com/watch?v=qYAr7EAlbOs (last accessed July 25, 2019) (watch from 1:16:44 to 1:17:21). *See also* Chi Almario-Gonzalez, Unmasking the trolls, Spin masters behind fake account, news sites, *available at* https://news.abs-cbn.com/focus/01/20/17/unmasking-the-trolls-spin-masters-behind-fake-accounts-news-sites (last accessed July 25, 2019).

499. *See* PHIL. CONST. art. XVI, § 11 (1).

500. Pål Storehaug, Social Media Marketing influence versus Mass Media, *available at* https://cloudnames.com/en/blog/social-media-marketing-influence (last accessed July 25, 2019).

thereof."[501] While social media is engaged in information dissemination, it cannot be considered mass media because, as mentioned earlier, the audience is both the creator and the audience; hence, the content made available online is not created by the social media company.[502] It simply acts a mode of transmitting the content created by the user to other users through the concept of social networking. This is more in line with the nature of social media and should not be considered as a mass media entity; thus, it should not be covered by the nationality requirement under the Constitution.

While this Article has established that social media companies should not be treated as mass media under Philippine law, it does not follow that they cannot be subject to state-imposed regulation.

Social media companies are treated by scholars and experts alike as online intermediaries. This refers to privately owned websites, servers, and routers[503] which provide a free and virtual soapbox where one may "regale the public."[504] It is through these internet intermediaries that content can be posted, shared, and distributed online. However, due to the large amount of information that is being published on their platforms, it becomes nigh impossible for these online intermediaries to review and correct every single content, unlike those of mass media companies such as print and radio, which have that opportunity.

While it can be argued that these online intermediaries have very little knowledge about the content published via their platform, it remains to be true that they are the most plausible subject matter for litigation as they are seen as the most effective point of control[505] over such content. Experts claim

501. Securities and Exchange Commission, Re: Marketing and Sale of Digital Publication Through the Internet and Mobile Technology; Advertising; Mass Media, SEC-OGC Opinion No. 14-06, at 2 (May 8, 2014) (citing Department of Justice, Opinion No. 40, Series of 1998 (Mar. 19, 1998)).

502. *See* Storehaug, *supra* note 500.

503. David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 377 (2010).

504. *Id.*

505. *See* Ardia, *supra* note 503, at 378.

that it is most strategic to go after the distribution network itself[506] considering that the online intermediaries have a commercial stake in these problems, which if not addressed, may lead to the company's downfall.

So how should the Philippines treat online intermediaries? It is first important to note that online intermediaries are not limited to social media companies but extend to search engines, auction sites, online forums, and other similar internet platforms which allows user interactions through the internet.[507]

Scholars further classified these intermediaries as either *active* or *passive* intermediaries. The distinction is important in determining the liability arising from the content. When an intermediary does not change the information published in its platform, it is considered *passive*.[508] If it does, it is deemed *active*.

As seen in the previous Section, states such as Germany have already legislated measures to impose liability on online intermediaries which do not take down a reported post within a specified period of time. While Germany's model is considered by many as unprecedented, its measure is akin to the *notice and take down* procedure discussed earlier in this Chapter.

It can be argued that platforms like Facebook actively manages content such as the trending section. In 2016, Facebook acknowledged that it employs human editors to pick and evaluate trending topics; hence, it meets the Council of Europe's definition of an editorial process.[509] Robert Thomson,

506. *See, e.g.*, New York v. Ferber, 458 U.S. 747, 759-60 (1982) & Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 364 (2014).

507. Mark MacCarthy, *What Payment Intermediaries are Doing about Online Liability and Why It Matters*, 25 BERKELEY TECH. L.J. 1037, 1038 (2010).

508. *See* Joint Declaration, *supra* note 427, ¶ 2 (a) & E.U. E-Commerce Directive, *supra* note 234, at 6.

509. Natali Helberger & Damian Trilling, Facebook is a news editor: the real issues to be concerned about (A Blog Post Published in the London School of Economics Media Policy Project blog), *available at* https://pure.uva.nl/ws/files/2738602/177517_Facebook_is_a_news_editor_the_real_issues_to_be_concerned_about.pdf (last accessed July 25, 2019).

Chief Executive of News Corporation, has stated that social media companies can no longer be considered as *passive* intermediaries,[510] to wit —

> 'These companies are in digital denial.' ... 'Of course, they are publishers and being a publisher comes with the responsibility to protect and project the provenance of news. The great papers have grappled with that sacred burden over decades and centuries, and you [cannot] absolve yourself from that burden or the costs of compliance by saying, [you are] are a technology company[ ].'[511]

Martin Sorell, Chief Executive Officer of WPP, the world's largest advertising company, claims that social media companies should be responsible for the content in their "digital pipes."[512] In reality, while the content posted on these platforms are not created by the companies, it is published and shared through it. They can no longer shy away from their obligation to ensure that their platforms are not used to spread fake news, most especially when it becomes a danger to public interest.

Mark Zuckerberg, Facebook's Chief Executive Officer, admitted that the company is a media company and not just a mere platform.[513] This admission is further supported by the additional mechanisms created by Facebook which includes "Facebook Live," which allows users to live stream.[514] These social media companies are no longer eligible for special protection considering the

510. *See* Richard Waters, et al., *Harsh truths about fake news for Facebook, Google and Twitter*, FIN. TIMES, Nov. 16, 2019, *available at* https://www.ft.com/content/2910a7a0-afd7-11e6-a37c-f4a01f1b0fa1 (last accessed July 25, 2019).

511. Waters, et al., *supra* note 510.

512. *Id.*

513 Matthew Ingram, *Mark Zuckerberg Finally Admits Facebook Is a Media Company*, FORTUNE, Dec. 23, 2016, *available at* http://fortune.com/2016/12/23/zuckerberg-media-company (last accessed July 25, 2019).

514. *See* Mathew Ingram, *Facebook's Claim That it Isn't a Media Company Is Getting Harder to Swallow*, FORTUNE, Dec. 15, 2016, *available at* https://fortune.com/2016/12/15/facebook-media-claim (last accessed July 25, 2019).

huge amounts of money they pool in every day by reason of the sheer number of users communicating through their platform.[515]

C. Closer Look on the Alternative: "Notice and Correct" Procedure

Throughout this Article, the two procedures — the *notice and correct* and the *notice and take down* have been extensively discussed. In this Section, their main difference is to be highlighted, and that is the role to be played by the social media company in the fake news problem.

The Authors agree that in a perfect world, the *notice and take down* procedure might be sufficient to address the problems arising from fake news published on social media. Imagine a situation where one person posts a fake news article online against A. A, being the affected party, can go to a competent court and ask the court to take down the assailed article. If the court agrees that the article is indeed false, the court can therefore ask the social media company to take it down from their platform. It is only upon the company's receipt of such order that said content will be removed.

While this may seem sufficient to address the fake news problem, the reality is that this model is not always possible considering that it can be quite burdensome for the courts to examine all applications for content removal given the large volume of fake news in the country. To impose this added responsibility to the courts — together with their existing duty to try and hear cases involving other forms of unprotected speech — may result to the further clogging of dockets and the prolonged resolution of cases.

Therefore, there arises a need for social media companies to play an active role in regulating content in their platform. As previously mentioned, online intermediaries play a passive role in relation to the content in their platforms. While these online intermediaries do not edit nor control the content published or shared by their users, these companies remain to have an economic interest over their platforms, which necessarily implies that these companies should also bear some responsibility. What this means is that these social media companies must either follow an order of a competent court

515. *See* Facebook Reports Fourth Quarter and Full Year 2016 Results, *available at* https://investor.fb.com/investor-news/press-release-details/2017/facebook-Reports-Fourth-Quarter-and-Full-Year-2016-Results/default.aspx (last accessed July 25, 2019).

mandating content removal, or after being notified, decide on whether or not the content must be deleted, subject to appeal to a judicial tribunal.

The latter is covered by the *notice and correct* procedure. Social media companies should have a similar obligation to *traditional* media companies to correct information once they have been notified of its alleged falsity or erroneousness. In the Philippines, Section 1 of the Philippine Journalist's Code of Ethics states that one has a duty "to correct substantive errors properly."[516] The Broadcast Code of 2011 also provides, under Article 5, that "when a mistake has been broadcast, it must be acknowledged and rectified as soon as possible by stating the mistake and making the correction."[517] While these currently do not apply to social media companies, the State must consider making these applicable to them, both in law and in principle. If these current press laws were to be applied to social media companies, considering that their effects are similar, if not greater, than traditional media, they should correct false news or information at the request of the real party in interest. If the platform decides to ignore the request, which it is entitled to do, the affected party would have the right to refer the case to a court or tribunal. Alternatively, this right to refer to the courts should also be given and guaranteed to the authors of the assailed content. They should be allowed to prove that there is nothing wrong with the subject matter.

The correction procedure should be akin to what is current applied in print media. An interested party who notifies the platform of the assailed content must receive a compulsory confirmation of receipt. Once the social media company is notified, it may choose any of the following actions: (1) do nothing and face the possibility of being brought to court; (2) perform fact-checking on the item and decide whether it should be deleted, corrected, or retained; (3) delete the item outright based on an in-house assessment; or (4)

516. Philippine Press Institute, Journalists' Code of Ethics, *available at* https://philpressinstitute.net/journalists-code-of-ethics-2 (last accessed July 25, 2019).

517. Kapisanan ng Mga Broadkaster ng Pilipinas, Broadcast Code of Ethics of the Philippines 2007 (as amended 2011) at 13, *available at* https://www.kbp.org.ph/wp-content/uploads/2008/04/KBP_Broadcast_Code_2011.pdf (last accessed July 25, 2019).

ask the author of the content, if he or she can correct the information, otherwise, it can face deletion.

Under this approach, the assailed content can be corrected either by the author or by the social media company itself. It is essential that this option be maintained. As mentioned above, many fake news stories are created and published with intent by agents who cannot be coerced into deleting them. In such cases, the social media company must take responsibility and intervene without the author's consent. Provided further that the social media company's technology is advanced and has the capacity to perform additional tasks, a follow-up obligation to distribute the corrected content to exactly the same audience as that which read the fake story in the first place can be imposed. This kind of obligation is akin to how in print media, it is required to publish the correction in the same medium, in the same place as the false information it has corrected.

To accomplish this measure, the Philippine Congress must enact legislation since Philippine laws are not so technologically advanced and did not put social media in consideration upon their formulation and enactment. Under this proposed law, social media companies must be vested with the obligation to be responsible for the content published on their respective platforms. However, the law will recognize that these social media companies do not have an obligation to monitor every single content being published on their platforms or to proactively search for false content, rather, their responsibility shall set in upon the notification made by the interested party as mentioned above. Otherwise, absent this notification requirement — as a condition *sine qua non* prior to the commencement of its responsibility — it may prove to be burdensome on the end of the social media company.

Nonetheless, social media companies should be given a reasonable amount of time to show that their self-imposed voluntary measures are sufficient to address the problem of fake news. A period of one to two years to determine if there is a further need for legislation is enough.

There exists a possibility that social media companies will be tempted to overreact and start taking down content upon notification, without trying to verify the claim, in order to avoid any form of liability. This creates an incentive on the part of the online intermediaries to censor a user's legitimate exercise of his right to freedom of expression in order to minimize exposure

from liability.[518] According to Felix Wu, more often than not, intermediaries would take down content upon receipt of a complaint[519] in relation to their "fragile commitment to the speech that they facilitate." [520] Actions made by online intermediaries insofar as the reported content is concerned will definitely give rise to issues on freedom of expression. These online intermediaries must be "neutral implementers of [ ] decisions[,]" not decision makers themselves."[521]

To minimize this risk, authors of social media content should be guaranteed the right to appeal against deletion.

IX. CONCLUSION AND RECOMMENDATIONS

What was once coined as the hope for modern democracy, social media has quickly transformed into a double-edged sword — in one side, the truth and in the other, falsity. True enough, social media has allowed every person, regardless of status, race, gender, and other qualification, to make his or her voice heard as long as he or she is connected to the Internet. This Article has emphasized time and again that social media has become so integrated to everyday life that it has given rise to numerous advantages and disadvantages. All of these are relevant in any democracy. Scholars say that regulation is not

518. Tess Marie P. Tan, Liberty and Prosperity in the Digital Age: Determining the Proper Treatment of Online Intermediaries in Light of the United Nations Guiding Principles on Business and Human Rights at 19-20, *available at* https://forlibertyandprosperity.files.wordpress.com/2018/03/flp-dissertation-contest-second-place.pdf (last accessed July 25, 2019) (citing Rebecca Ong, *Internet Intermediaries: The Liability for Defamatory Postings in China and Hong Kong*, 29 COMPUTER L. & SEC. REV. 274, 281 (2013) & Daithí Mac Síthigh, *The fragmentation of intermediary liability in the UK*, 8 J. INTELL. PROP. L. & PRAC. 521, 525-26 (2013)).

519. *See* Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Liability*, 87 NORTE DAME L. REV. 293, 301 (2011)

520. Wu, *supra* note 519, at 307 n. 64 (citing Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 28 (2006)).

521. Marcelo Thompson, *Beyond Gatekeeping: The Normative Responsibility of Internet Intermediaries*, 18 VAND. J. ENT. & TECH. L. 783, 785 (2016).

the solution; rather, government institutions should exert more effort to educate their constituents as to the importance of the truth.[522]

This Article has shown that scholars in favor of freedom of expression will relate any attempt to regulate social media as government censorship. In fact, there is some agreement to the proposition that ideas, regardless of their gravity, should be allowed to be published and that the burden to determine what is right from wrong should be with the users themselves.[523]

While this argument may seem appealing, the harsh reality is that this is not what is happening in the present day. If an international actor decides to create an intentionally fake story, does that constitute as a genuine reflection of the user's thoughts? Should the government just stand idly knowing that so many of its people are relying and basing their opinion about genuine issues on false information? In fact, advocates of regulating social media content argue that the psychological and societal effects brought by the proliferation of false information has been so damning that the need for state action has become imperative. This Article goes so far in contending that if the State allows the perpetuation of these fake news online, it can violate both the due process and equal protection clauses of the Constitution.

The Economist made a conclusion that the era of digital exceptionalism cannot last forever because the power of these platforms' over public life, economy, and the international community is becoming so great. [524] Overwhelming evidence shows that these social media companies' business models and construction unintentionally entrench echo chambers.[525] Clearly, the best long term solution to combat fake news is to educate and improve the social media literacy of the people and help them identify what is true from what is not. Unfortunately, this will take several years until the effects of such

522. Žiga Turk, Why a Crackdown on Fake News is a Bad Idea, *available at* http://www.thedigitalpost.eu/2017/channel-innovation/why-a-crackdown-on-internet-fake-news-is-a-bad-idea (last accessed July 25, 2019).

523. *See, e.g.*, John Samples, Why the Government Should Not Regulate Content Moderation of Social Media, *available at* https://www.cato.org/publications/policy-analysis/why-government-should-not-regulate-content-moderation-social-media#full (last accessed July 25, 2019).

524. *Eroding Exceptionalism: Internet firms' legal immunity is under threat*, *supra* note 229.

525. Tufekci, *supra* note 110.

educational campaigns will set in. Hence, there is a need to impose both a long-term and quick-fix solution to address the problems this country currently faces.

Decisive action is imperative in solving this issue. The proliferation of fake news heavily influences public discourse and must be quickly stopped. Democracy is in danger because of its effects. Fake news is no longer just a joke — it is a threat as real and as dangerous as hate speech, fighting words, libel, and slander.

True enough, there are existing measures that aim to fight fake news such as the increasing number of fact-checkers and anti-fake news software or blockers. However, these are insufficient. As explained in earlier Chapters, existing press laws appear to be the best foundation in formulating an approach to solve this problem. It must be remembered that excessive restriction of free speech should be highly frowned upon regardless of the danger fake news poses.[526] It is the position of the Authors that social media companies should not be allowed to arbitrarily decide what is the truth. There is obviously a gap in the current structure of governance of content available online. However, the solution should not be skewed in favor of resolving what can be said online on the hands of these private corporations. It must be remembered, as earlier pointed out, that these social media companies do not operate primarily for public interest, but for commercial and financial purposes. It is precisely because of this that the State must impose minimum standards to fill in these gaps such as ensuring that the fact-checking process is performed by humans and not by artificial intelligence. While this may result to heavier costs, these social media companies must adjust to the demands of the State considering that public interest weighs far greater than their corporate agenda.

It therefore becomes clear that both active and passive intermediaries have the same rights to publish and impart information, as well as the obligation to protect the freedom of expression.

The ruling in *eBay* illustrates the exception to the general rule. It can therefore be argued that all intermediaries are expected to, to a certain degree, curate the content they publish, or at the very least, install safety mechanisms

526. The Facebook 'fake news' scandal is important – but regulation isn't the answer, *available at* https://www.independent.co.uk/voices/editorials/the-facebook-fake-news-scandal-is-important-but-regulation-isnt-the-answer-a7419386.html (last accessed July 25, 2019).

so their infrastructures are not abused in order to violate individual liberties. However, it should not be made to undertake censorship on behalf of the State.

Thus, pursuant to both the Constitution as well as the country's international obligations to protect and promote the right to freedom of expression, it must be observed that:

(1) As a general rule, social media companies are immune from liability for the fake news published and/or disseminated in their platform, provided that they have no involvement in such content;

(2) However, social media companies shall be liable if, after the notice and correct procedure have been complied with, they still fail to act on the assailed content identified by the complainant.

(3) Any law which seeks to regulate individuals from publishing fake news is most likely to be struck down as unconstitutional due to the lack of adequate and sufficient standards provided by law as regards the determination of what is true and what is not. Any law that seeks to do so shall cause a chilling effect to the freedom of expression, which is unconstitutional.

(4) The immunity afforded to social media companies is not absolute because they are not expected to be mere bystanders to unlawful content published and disseminated on their platforms. Although they are not expected to act and serve as a government censor, the burden of scrutinizing content cannot be entirely lodged to the judicial courts for that would be too costly and burdensome. It is therefore more effective if interested parties be allowed to give notice to the social media company regarding an assailed false content and let the notice and correct procedure take place, and only if the social media company is shown to have not exerted the appropriate mechanisms should it be held liable, subject to an action filed before an independent and competent tribunal. Further, the absence of editorial control is arguably immaterial considering the economic benefit that the social media company derives from the content engaged upon in its platform.

In view of the foregoing, the Authors hereby recommend, to wit —

Social media companies must be given a prescriptive period, within the period of two years, to affirm to the Philippine Government that they can both effectively and efficiently combat fake news on their own. On the other hand, the Philippine Government, together with non-governmental organizations, civil society representatives, and the media sector, shall evaluate the existing voluntary measures of these social media companies. If the evaluation results in a negative assessment, the Philippines shall impose a top-down regulation;

In the event that the State is unsatisfied with the voluntary self-regulatory measures of the social media company, it shall, through Congress, pass a law which shall impose liability on these social media companies.

This law shall embody and introduce both the *notice and correct* and *notice and take down* procedures, with a strong emphasis on the former.

The *notice and correct* procedure shall be implemented accordingly, thus —

Any party, whether natural or juridical, is allowed to inform, notify, or report to the social media company of the assailed false and/or harmful content, provided that his or her interests are prejudiced by such content;

The social media company must then provide mandatory confirmation and acknowledgment of the report made by the interested party;

In turn, the social media company is allowed to do any of the following:

(1) Do nothing;

(2) Delete the content; or

(3) Refer the assailed content to either in-house or independent fact-checkers, and based on their assessment, the social media company shall make a decision as to whether or not to retain the content, correct it, ask the author to correct it or else it will be deleted, or delete it entirely.

The moment the social media company decides to delete or correct the assailed content, the authors of the content must be notified of the reasons justifying the action. Further, the authors need to be given the right to appeal to an independent body, preferably a court or an arbitration tribunal, if they feel and consider that there has been a violation of their rights because of the social media company's action or inaction.

Through this measure, there is no violation of the constitutional right to freedom of speech and expression of the author. The one who published the

content on social media is not prevented from posting his or her content online. If the content he or she posted was reported by another user who believes that his or her interests were prejudiced by said content, he or she is not left without any recourse because the proposed law must give him or her the opportunity to appeal the decision of the social media company as to the treatment of his or her content. With all these taken into consideration, the constitutional challenges identified in this Article are duly addressed.