Behavioural Compliance and Law Enforcement in Online Hate Speech

Pompeu CASANOVAS [1, 2, 3] and Andre OBOLER [1, 2, 4]

[1] *La Trobe Law School, La Trobe University, Melbourne, Australia*
[2] *Data to Decisions Cooperative Research Centre*
[3] *Autonomous University of Barcelona (IDT), Spain*
[4] *Online Hate Prevention Institute (OHPI), Australia*

Abstract. It is usually said that technical solutions should operate ethically, in compliance with the law and subject to good governance principles. In this position paper we face the problem of behavioural compliance and law enforcement in the case of hate and fear speech online. Law enforcement and behavioural compliance are ways of coping with the objective of stopping hate online. We contend that a combination of regulatory instruments, incentives, training, proactive self-awareness and education can be effective to create legal ecosystems to improve the present situation.

Keywords. Hate speech, rule of law, semantics, NLP, legal governance,

CEUR-WS.org/Vol-2309/11.pdf

1. Introduction

Violence is a pervasive phenomenon in contemporary global societies. It has been fostered by the expansion of the Internet, social media networks and the fast development of the web of data. Violent language reflected in bias attitudes is the first step in the pyramids of hate and escalation of conflicts. Even in the most extreme case of inhumanity, Rabbi Abraham Joshua Heschel noted how “the Holocaust did not begin with the building of crematoria, with tanks and guns. It began with uttering evil words, with defamation, with language and propaganda” [1]. This is the opinion of most linguists in the 20th and 21st c., e.g. [2]. According to some recent studies, media and the way in which minority groups are targeted are fuelling this phenomenon. Dichotomic, binary categories, and the practice of depicting non-white cultures as “alien” (“othering”), play a major role in reinforcing negative, weak, or fearful images of migrants and refugees and spreading xenophobia [3].

However, detecting, tracking, and monitoring these particular uses of language on the web has turned out to be a difficult task, as it implies a meta-cognitive operation of annotating, classifying and clustering terms and expressions from a previous interpretation of their context of usage. Hate speech can partially be fear speech as well. But, what is hate and what is fear disguised by hate speech? [4] Violence attracts, fascinates and repeal, as shown by the 'beautiful' war images displayed newspapers and on the media [5].

In this position paper, we contend that (i) it is much better to take a proactive ethical stance than adopting a passive *laissez-faire* approach, (ii) there is an effective possibility of making errors of judgment (false positives and negatives), (iii) technology offers at present some means to overcome or at least reduce these risks (although not completely), (iv) the rise of online hate speech is an indicator of cultural change that should be taken seriously, (v) there is no simple solution to stop this based on traditional legal instruments (i.e. enactment of rules and enforcement of laws), (vi) hence, some regulatory imagination is needed, stemming from a combination of hard and soft law, smart regulations, multi-stakeholder governance, policies and ethics.

2. Definition

The first problem is the meaning of the expression. We can identify three stages: (i) before World War II and in the inter-war period hate speech was defined as 'race hate' or 'group libel', (ii) in the second half of the past century, definitions become more inclusive and sensitive to victimisation processes, e.g. Human Rights Watch defined it as 'any form of expression regarded as offensive to racial, ethnic and religious groups and other discrete minorities, and to women' (iii) in the 21st century, even this meaning that included all kind of sexual and political biases has been broadened to cover all kinds of oppression (religious, cultural, political or technological —i.e. based on the lack of knowledge or technological skills) [6]. The idea is that human rights and its political side, civil rights, are deemed to *empower* people; hence, all sorts of humiliation implies a loss of dignity that constitutes in itself a form of *disempowerment*, i.e. an aggression that can be qualified as a form of violence. Violence finds its own 'connectomes' on the Internet, producing a permanent and structural harm that can be easily amplified for political and economic reasons [7]. Words create worlds, that is, they shape the very fabric of our environment. In "linked democracy" scenarios, this particular threat should be avoided and considered the first step to tyranny, thus, a negative condition for the construction of the global (linked) space [8].

This approach represents a turning point that shifts the way in which the jurisprudence and legal philosophy of the 20th c. described the problem as a constituent of political democracies. The USA is the only Western democracy to exclude any kind of legal punishment against extreme forms of language intended to foster hatred in the public space.[1] Free speech, the First Amendment provision, prevails. Against hate speech

[1] The *International Convention on the Elimination of All Forms of Racial Discrimination* entered into force on January 4th 1969 [28]. It has been ratified by 88 states. The Convention also requires its parties to outlaw hate speech and criminalize membership in racist organizations. USA ratified the Convention, but upon ratification, it stated the following reservations: "1.That the Constitution and laws of the United States contain extensive protections of individual freedom of speech,

bans one of the more persuasive arguments was advanced by Ronald Dworkin [9], who pointed out that law enforcement would deny subjects an adequate opportunity for dissent. Freedom of speech 'guarantees and preserves liberalism's commitment to equality by offering everyone an opportunity to speak, whereas any other policy, such as state regulation, would fail to offer this equal opportunity' [10]. This egalitarian liberalism has recently been contested by Jeremy Waldron, stemming from [11] the perspective of the construction of a public space based on dignity, a human constituent that cannot be politically bartered nor negotiated.

3. Technology: fostering dignity

From a technological point of view the nature of the argument, fostering dignity, has been perceived as a real need:

> The exponential growth in the Internet as a means of communication has been emulated by an increase in far-right and extremist web sites and hate based activity in cyberspace. The anonymity and mobility afforded by the Internet has made harassment and expressions of hate effortless in a landscape that is abstract and beyond the realms of traditional law enforcement. This paper examines the complexities of regulating hate speech on the Internet through legal and technological frameworks. It explores the limitations of unilateral national content legislation and the difficulties inherent in multilateral efforts to regulate the Internet. [12]

> In the realms of social media, hate speech is a kind of writing that disparages and is likely to cause harm or danger to the victim. It is a bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics [6]. It is a kind of speech that demonstrates a clear intention to be hurtful, to incite harm, or to promote hatred. The environment of social media and the interactive Web 2.0 provides a particularly fertile ground for creation, sharing and exchange of hate messages against a perceived enemy group. These sentiments are expressed at news review sites, Internet forums, discussion groups as well as in micro-blogging sites. [13]

> We address the problem of hate speech detection in online user comments. Hate speech, defined as an abusive speech targeting specific group characteristics, such as ethnicity,

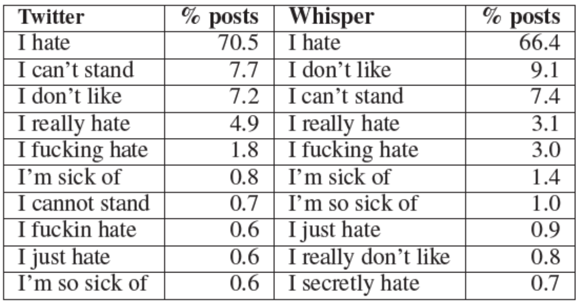
expression and association. Accordingly, the United States does not accept any obligation under this Convention, in particular under articles 4 and 7, to restrict those rights, through the adoption of legislation or any other measures, to the extent that they are protected by the Constitution and laws of the United States. 2. That the Constitution and laws of the United States establish extensive protections against discrimination, reaching significant areas of non-governmental activity. Individual privacy and freedom from governmental interference in private conduct, however, are also recognized as among the fundamental values which shape our free and democratic society. […] 3. That with reference to article 22 of the Convention, before any dispute to which the United States is a party may be submitted to the jurisdiction of the International Court of Justice under this article, the specific consent of the United States is required in each case." Cfr. https://treaties.un.org/Pages/ViewDetails.aspx?src=TREATY&mtdsg_no=IV-2&chapter=4&lang=en#EndDec

> religion, or gender, is an important problem plaguing websites that allow users to leave feedback, having a negative impact on their online business and overall user experience. [14]

Automated detection, clustering, monitoring and managing, and tracking on real time are the most common problems. Several approaches have been proposed so far, mostly leaning on NLP, AI and semantics: (i) classifiers can be used to detect the presence of hate speech, using sentiment analysis and subjectivity detection in pre-defined areas (e.g. race, gender, religion) [35], (ii) lexicons can be created and also used for this purpose, (iii) practical projections to real-world discourses can then be applied [16], (iv) distributed low-dimensional representations of hate comments can be identified using neural language models that can then be fed as inputs to a classification algorithm [14], (v) machine learning [15], (vi) annotated datasets, impact of extra-linguistic features in conjunction with character n-grams for hate speech detection [16] [17], (vii) qualitative and discourse analysis [16]. The table below displays the top ten expressions in Twitter and Wisper [19].

A recent survey on NLP methods also furnishes several examples [20]:

> (1) Go fucking kill yourself and die already useless ugly
> pile of shit scumbag.
> (2) The Jew Faggot Behind The Financial Collapse
> (3) Hope one of those bitches falls over and breaks her leg



Twitter	% posts	Whisper	% posts
I hate	70.5	I hate	66.4
I can’t stand	7.7	I don’t like	9.1
I don’t like	7.2	I can’t stand	7.4
I really hate	4.9	I really hate	3.1
I fucking hate	1.8	I fucking hate	3.0
I’m sick of	0.8	I’m sick of	1.4
I cannot stand	0.7	I’m so sick of	1.0
I fuckin hate	0.6	I just hate	0.9
I just hate	0.6	I really don’t like	0.8
I’m so sick of	0.6	I secretly hate	0.7

While the set of features examined by [20] in the different works present a great diversity, the classification methods mainly focus on supervised learning, surface-level features to classify, and generic features, such as bag of words or embeddings. According to the authors, character-level approaches work better than token-level approaches, and lexical resources, such as list of slurs, may help classification, but usually only in combination with other types of features. A benchmark or annotated dataset would be needed, as inferences, suppositions and associative tropes are difficult to detect and could benefit from a semantic approach considering the contexts and possible scenarios.

An interesting approach is taken when annotations and descriptions are ground on a crowdsourced-bases. Oboler [21] identified ten years ago the main elements of antisemitic discourse in social media —what he called “antisemitism 2.0”— as follows: (i) The content denies its antisemitic nature; (ii) it promotes antisemitic tropes , (iii) it claims its message is a legitimate view people should be free to hold (no different from choosing to support a particular sports team), (iv) the content is designed to go viral by making sharing the content both technically easy and socially acceptable in social media, (v) the audience is not the dedicated antisemites but rather the susceptible public.

The next stage has been the creation of social and collective bonds, seeking for awareness and participation [22] [23]:

> Based on the recommendations of the Global Forum, the Online Hate Prevention Institute (OHPI) in Australia developed FightAgainstHate.com, a cloud based tool for reporting, monitoring, and measuring the response to online antisemitism as well as other forms of online hate. Using the tool the public can report various types of online hate speech and assign both a category and sub-category to the hate they report.

4. Regulatory models: socio-legal ecosystems

How should hate speech be effectively regulated? How can compliance with universal values such as peace and tolerance be achieved?

Banks [12] suggests that "a broad coalition of government, business and citizenry is likely to be most effective in reducing the harm caused by hate speech".

This is a reasonable goal, but not easily achievable. Some governments can use hate speech for other political reasons —e.g. to prosecute citizens participating in demonstrations.

We think that what is required is a set of regulatory tools to create *socio-legal ecosystems*, e.g. patterns of behaviour able to show resilience, i.e. *leaning on behavioural rather than normative compliance* [7] [24] Even though, this is not simple.

Behavioural compliance has been investigated in organisations, companies, and administrations. Several studies highlight the importance of social bonding, social influence, and cognitive processing [25] [26]. Deterrence does not suffice [27]. Social bonds largely influence attitudes toward compliance and foster the adoption of personal codes of conduct. However, social bonds that work against racism are not spontaneous. Waseem [17] concludes:

> We find that amateur annotators are more likely than expert annotators to label items as hate speech, and that systems trained on expert annotations outperform systems trained on amateur annotations.

Thus, expert knowledge, guidance (and political will), matter [28]. To make effective the protections of the rule of law in the age of linked data, a combination of sanctions, training, and educative efforts should be put in place. Therefore, ethics should play a new regulatory role on the web of data. We prefer the expression "legal governance" rather than "law". This is a new cultural turn not (or not only) for coercive measures, but for *relational law and justice* on the web of data [29].

5. Final remarks: behavioural compliance

We would like to rise some more questions to shed some light on this debate. Behavioural compliance is more difficult to achieve than regulatory compliance, for more conditions apply to the available regulatory means and instruments. Enforcement can only be a component, along with agreement, conformance, and acceptance of values, principles and rules. Hence, the acquiescence and cooperation of the subjects must be represented as a necessary condition for the regulatory pattern to occur.

Therefore, the tension between free speech and hate speech limitations cannot be solved in one single dimension. At the epistemic level we should introduce (i) the complexity entailed by collective interactions and decision-making, (ii) the different levels of abstraction in which these concepts are used, (iii) the micro- and macro- societal layers in which the implementation of regulations operate.

Gould observes that 'hate speech is fuzzed in the abstract but more apparent when confronted in person' [31]. He carried out an interesting empirical analysis, showing that despite the judicial hurdles based on the first amendment the concept has pervaded American society. We are not facing a discrete category, but a continuum in which semantic and pragmatic elements are entangled to produce social adhesion and bonds. This would be an example of *societal regulation:*

> Hate speech regulation has permeated other elite institutions like the media and has trickled down to influence mass opinion and common understandings of institutional norms. [So] extra-judicial law and the power of legal meaning-making […] informal law or mass constitutionalism is as powerful as the formal constitution, providing vehicles to change that exists without the intervention of courts. [33]

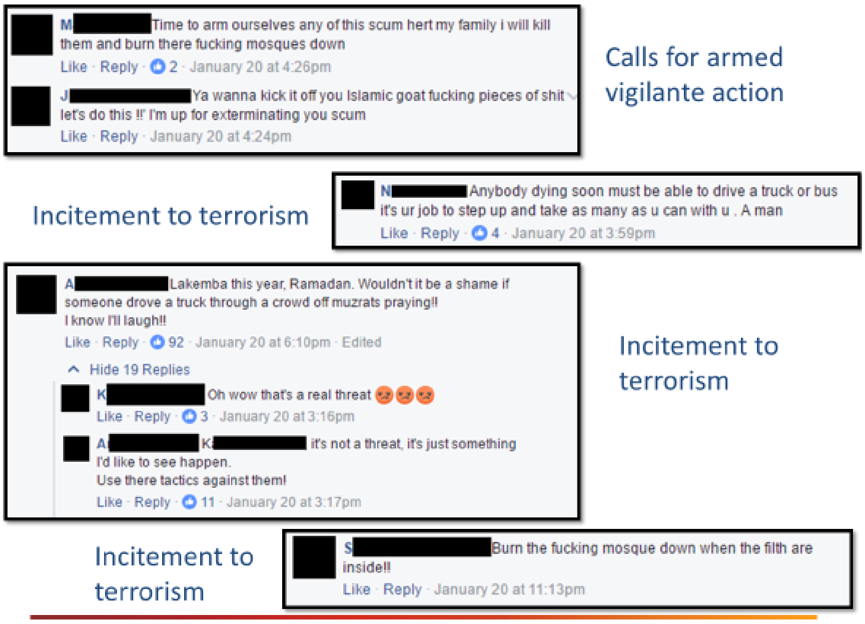
Delgado and Stefancic [32] observe that, at least in USA, there is a tendency to frame the debate in "legal" terms, i.e. as one of procedure rather than substance. On the contrary, defenders of setting hate speech limitations: (i) ponder the importance of social power, and recognize the connection between general, nontargeted hate speech and the rise of destructive social movements, (ii) point out that hate speech often targets individuals who, by reason of his or her race or physical appearance, have been the object of similar attacks many times before.

Reliability of annotations raise another problem, as "the presence of hate speech should perhaps not be considered a binary yes-or-no decision, and raters need more detailed instructions for the annotation." [33] Researchers working on a German hate speech corpus for the refugee crisis in 2016 noticed that building a classifier (i.e. rating the offensives of tweets on a 6-point Likert scale) entailed discussions not only among raters but researchers, due to personal attitudes.

The difficulty of automated detections should not be underestimated. In the recent First Shared task on Aggression Identification organized with the TRAC workshop at COLING 2018, in which 30 teams finally submitted their system, "performance of the neural networks-based systems as well as the other approaches do not seem to differ much. If the features are carefully selected, then classifiers like SVM and even random forest and logistic regression perform at par with deep neural networks" [34]. The task was to develop a classifier that could discriminate between Overtly Aggressive, Covertly Aggressive, and Non-aggressive texts. The participants were provided with a dataset of 15.000 aggression-annotated posts and comments (in English and Hindi). Systems obtained a weighted F-score between 0.50 and 0.64. This is consistent with similar scores in current researches summarised in this paper (section 3).

Thus, crowdsourced hate speech reporting face two main challenges: (i) cooperation between lay and expert knowledge to annotate the corpus, (ii) the difference between the surface of discourse and the environments and contexts that discourses contribute to create.

What is crucial is differentiating between the individual expression and the course of collective action in which this expression is embedded. This would help to separate hate speech from fear speech. Figures 1 and 2 show how cooperation between lay people (reporting), experts (evaluating and counselling) and institutions (receivers) can help to solve the puzzle. But even in this case, independent monitoring and evaluation matters, as governments may fail in reducing the volume of abusive content on social media corporations [36]. In addition, some governments may also divert the definition of hate speech, broadening it to target political adversaries. Thus, hate speech regulations should not be understood only from a narrow national perspective, but as a global exercise of implementation of human and democratic rights.



[redacted] [redacted] Time to arm ourselves any of this scum hert my family i will kill them and burn there fucking mosques down
Like · Reply · 👍 2 · January 20 at 4:26pm

[redacted] [redacted] Ya wanna kick it off you Islamic goat fucking pieces of shit let's do this ‼ I'm up for exterminating you scum
Like · Reply · January 20 at 4:24pm

Calls for armed vigilante action

Incitement to terrorism

[redacted] [redacted] Anybody dying soon must be able to drive a truck or bus it's ur job to step up and take as many as u can with u . A man
Like · Reply · 👍 4 · January 20 at 3:59pm

[redacted] [redacted] Lakemba this year, Ramadan. Wouldn't it be a shame if someone drove a truck through a crowd off muzrats praying!!
I know I'll laugh!!
Like · Reply · 👍 92 · January 20 at 6:10pm · Edited

^ Hide 19 Replies

[redacted] [redacted] Oh wow that's a real threat 😡😡😡
Like · Reply · 👍 3 · January 20 at 3:16pm

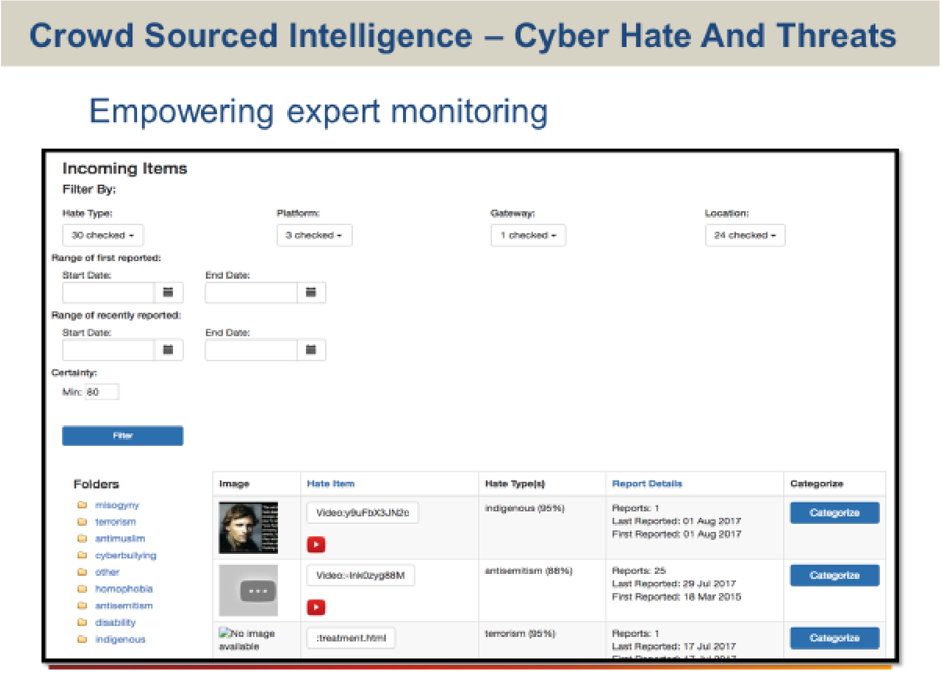
[redacted] [redacted] [redacted] it's not a threat, it's just something I'd like to see happen.
Use there tactics against them!
Like · Reply · 👍 11 · January 20 at 3:17pm

Incitement to terrorism

Incitement to terrorism

[redacted] [redacted] Burn the fucking mosque down when the filth are inside!!
Like · Reply · January 20 at 11:13pm

Fig. 1. Types of organised threats. Source: Oboler [23]



Crowd Sourced Intelligence – Cyber Hate And Threats

Empowering expert monitoring

Incoming Items

Filter By:

Hate Type: 30 checked

Platform: 3 checked

Gateway: 1 checked

Location: 24 checked

Range of first reported:

Start Date:

End Date:

Range of recently reported:

Start Date:

End Date:

Certainty:

Min: 80

Folders

- misogyny
- terrorism
- antimuslim
- cyberbullying
- other
- homophobia
- antisemitism
- disability
- indigenous

Image	Hate Item	Hate Type(s)	Report Details	Categorize
	Video:y9uFbX3JN2c	indigenous (95%)	Reports: 1 Last Reported: 01 Aug 2017 First Reported: 01 Aug 2017	Categorize
	Video:-Ink0zyg88M	antisemitism (88%)	Reports: 25 Last Reported: 29 Jul 2017 First Reported: 18 Mar 2015	Categorize
No image available	:treatment.html	terrorism (95%)	Reports: 1 Last Reported: 17 Jul 2017	Categorize

Figure 2. Facilitation of experts' tasks. Source: Oboler [23]

Acknowledgements

This research was partially funded by the Data to Decisions Cooperative Research Centre (D2D CRC, Australia) [2], and Meta-Rule of Law (DER2016-78108-P, Spain). Views expressed herein are however not necessarily representative of the views held by the funders.

References

[1] Eisen, A. The Spiritual Audacity of Abraham Joshua Heschel, *On Being*, 6 December 2012 https://onbeing.org/programs/arnold-eisen-the-spiritual-audacity-of-abraham-joshua-heschel/

[2] Klemperer, V. LTI. Lingua Tertii Imperii. A Philologist's Notebook [1975], London. Continuum (2006)

[3] Naffi, N. The Trump effect in Canada: A 600 per cent increase in online hate speech. November 2, 9.37am AEDT The Conversation. (2017) https://theconversation.com/the-trump-effect-in-canada-a-600-per-cent-increase-in-online-hate-speech-86026

[4] Naffi, N. Ceci n'est pas un discours haineux, 23/04/2017 08:21 EDT | Actualisé 23/04/2017 08:21 EDT https://quebec.huffingtonpost.ca/nadia-naffi/islamophobie-discours-haineux_b_16151548.html

[2] http://www.d2dcrc.com.au/

[5] Shields, D. War Is Beautiful. The New York Times Pictorial Guide to the Glamour of Armed Conflict. Nova York: Powerhouse Books (2015)

[6] Siegel, M.L. Hate speech, civil rights, and the Internet: The jurisdictional and human rights nightmare. Alb. LJ Sci. & Tech., 9, p.375-398 (1998)

[7] Poblet, M., Casanovas, P., Rodríguez-Doncel, V. Linked Democracy. Cham, Springer Briefs (2019)

[8] Casanovas, P., Mendelson, D. and Poblet, M., A Linked Democracy Approach for Regulating Public Health Data. Health and Technology, 7(4), pp.519-537 (2017)

[9] Dworkin, R. Freedom's Law. Oxford, Oxford University Press (1996)

[10] Levin, A. Pornography, Hate Speech, and Their Challenge to Dworkin's Egalitarian Liberalism. Public Affairs Quarterly, Vol. 23, No. 4, pp. 357-373 (2009)

[11] Waldron, J. The Harm in Hate Speech, MA, Cambridge University Press (2012)

[12] Banks , J. Regulating hate speech online, International Review of Law, Computers & Technology, 24:3, 233-239, DOI: 10.1080/13600869.2010.522323 (2010)

[13] Gitari, N.D., Zuping, Z., Damien, H. and Long, J., A lexicon-based approach for hate speech detection. International Journal of Multimedia and Ubiquitous Engineering, 10(4), pp.215-230. (2015)

[14] Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V. and Bhamidipati, N., 2015, May. Hate speech detection with comment embeddings. In Proceedings of the 24th international conference on world wide web (pp. 29-30). ACM (2015)

[15] Burnap, P., Williams, M.L.. Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. Policy & Internet, 7(2), pp.223-242 (2015)

[16] Waseem, Z. and Hovy, D. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. Proceedings of NAACL-HLT 2016, pages 88–93, San Diego, California, June 12-17, Association for Computational Linguistics (2016).

[17] Waseem, Z.. Are you a racist or am I seeing things? Annotator influence on hate speech detection on Twitter. In Proceedings of the first workshop on NLP and computational social science (pp. 138-142) (2016).

[18] Erjavec, K., Kovačič, M.P., 2012. “You Don't Understand, This is a New War!” Analysis of Hate Speech in News Web Sites' Comments. Mass Communication and Society, 15(6), pp.899-920 (2012)

[19] Silva, L.A., Mondal, M., Correa, D., Benevenuto, F. and Weber, I.. Analyzing the Targets of Hate in Online Social Media. In ICWSM (pp. 687-690) (2016)

[20] Schmidt, A. and Wiegand, M. A survey on hate speech detection using natural language processing. In Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media (pp. 1-10) Valencia, Spain, April 3-7, 2017 Association for Computational Linguistics (2017)

[21] Oboler, Online Antisemitism 2.0. “Social Antisemitism” on the “Social Web”, JCPA, 1 April, (Pre-released in February 2008)

[22] Oboler, A. Measuring the Hate. The State of Antisemitism in Social Media. Online Hate Prevention Institute. Produced for the Global Forum for Combating Antisemitism (2016)

[23] Oboler, A. Building peace by fighting online hate. Yitzhak Rabin Memorial Lecture, 4 November 2018, slides. (2018)

[24] Gunderson L. & Cosens B. Case Studies in Adaptation and Transformation of Ecosystems, Legal Systems, and Governance Systems. In Cosens B., Gunderson L. (eds) Practical Panarchy for Adaptive Water Governance. Springer: Cham. (2018)

[25] Ifinedo, P. Information systems security policy compliance: An empirical study of the effects of socialisation, influence, and cognition. Information & Management, 51(1), pp.69-79 (2014)

[26] Vroom, C. and Von Solms, R., 2004. Towards information security behavioural compliance. Computers & Security, 23(3), pp.191-198. (2004)

[27] Ogbonna, E. Harris, L.C.Managing organizational culture: compliance or genuine change?. British Journal of Management, 9(4), pp.273-288 (1998)

[28] Oboler, A. Technology and regulation must work in concert to combat hate speech on line. March 12, 2018 6.09pm AEDT (2018) https://theconversation.com/technology-and-regulation-must-work-in-concert-to-combat-hate-speech-online-93072

[29] Casanovas, P., Poblet, M. Concepts and fields of relational justice. In Computable Models of the Law (pp. 323-339). LNAI 4884, Springer, Berlin, Heidelberg (2008)

[30] United Nations. International Convention on the Elimination of All Forms of Racial DiscriminationAdopted and opened for signature and ratification by General Assembly resolution 2106 (XX) of 21 December 1965. https://www.ohchr.org/en/professionalinterest/pages/cerd.aspx

[31] Gould, J.B. Speak no evil: The triumph of hate speech regulation. University of Chicago Press (2010)

[32] Delgado, R.., Stefancic, J., Four observations about hate speech. Wake Forest L. Rev., 44, p.353-370 (2009)

[33] Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N. and Wojatzki, M., Measuring the reliability of hate speech annotations: The case of the european refugee crisis. arXiv preprint arXiv:1701.08118. (2017)

[34] Kumar, R., Ojha, A.K., Malmasi, S. and Zampieri, M., Benchmarking Aggression Identification in Social Media. In Proceedings of the First Workshop on Trolling, Aggression and Cyberbullying (TRAC-2018) pp. 1-11 Santa Fe, USA, August 25 (2018)

[35] Gambäck, B. and Sikdar, U.K., 2017. Using convolutional neural networks to classify hate-speech. In Proceedings of the First Workshop on Abusive Language Online (pp. 85-90).

[36] Oboler, A. Technology and regulation must work in concert to combat hate speech online. March 12, 2018 6.09pm AEDT. https://theconversation.com/technology-and-regulation-must-work-in-concert-to-combat-hate-speech-online-93072