



Universiteit
Leiden
The Netherlands

Algorithmic speech and freedom of expression

Sears, A.M.

Citation

Sears, A. M. (2020). Algorithmic speech and freedom of expression. *Vanderbilt Journal Of Transnational Law*, 53(4), 1327-1376. Retrieved from <https://hdl.handle.net/1887/139144>

Version: Publisher's Version

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/139144>

Note: To cite this publication please use the final published version (if applicable).

Algorithmic Speech and Freedom of Expression

Alan M. Sears*

ABSTRACT

Algorithms have become increasingly common, and with this development, so have algorithms that approximate human speech. This has introduced new issues with which courts and legislators will have to grapple. Courts in the United States have found that search engine results are a form of speech that is protected by the Constitution, and cases in Europe concerning liability for autocomplete suggestions have led to varied results. Beyond these instances, insight into how courts handle algorithmic speech are few and far between.

By focusing on three categories of algorithmic speech, defined as curated production, interactive/responsive production, and semi-autonomous production, this Article analyzes these various forms of algorithmic speech within the international framework for freedom of expression. After a brief introduction of that framework and a look towards approaches to algorithmic speech in the United States, the Article then examines whether the creators or controllers of different forms of algorithms should be considered content providers or mere intermediaries, the determination of which ultimately has implications for liability, which is also explored. The Article then looks at possible interferences with algorithmic speech, and how such interferences may be examined under the three-part test—particular attention is paid to the balancing of rights and interests at play—in order to answer the question of the extent to which algorithmic speech is worthy of protection under international standards of freedom of expression. Finally, other relevant issues surrounding algorithmic speech are discussed that will have an impact going forward, many of which involve questions of policy and societal values that accompany granting algorithmic speech protection.

* Researcher and Lecturer at eLaw, Center for Law and Digital Technologies, Faculty of Law, Leiden University, The Netherlands. LL.M., Leiden Law School (2017); J.D., Notre Dame Law School (2014); B.A., Baylor University (2006). The author would like to thank Professor Jan Oster for his insightful comments on earlier drafts. Contact: a.m.sears@law.leidenuniv.nl

TABLE OF CONTENTS

I.	INTRODUCTION	1328
	A. <i>Background</i>	1328
	B. <i>What is Algorithmic Speech?</i>	1331
II.	ALGORITHMIC SPEECH AND THE SUBSTANTIVE SCOPE OF FREEDOM OF EXPRESSION.....	1335
	A. <i>Frameworks for Freedom of Expression</i>	1335
	1. The International Framework.....	1335
	2. The United States' Framework and Algorithmic Speech	1337
	B. <i>How Might Algorithmic Speech Fit into the International Framework</i>	1341
	1. Attribution of Algorithmic Speech and Status as a Content Provider or Intermediary	1341
	2. Liability for Algorithmic Speech	1349
	a. The European Framework	1350
	b. The United States and Elsewhere	1356
	c. Limited Liability Generally.....	1359
III.	TO WHAT EXTENT IS ALGORITHMIC SPEECH WORTHY OF PROTECTION UNDER INTERNATIONAL STANDARDS OF FREEDOM OF EXPRESSION?.....	1360
	A. <i>Interferences with Algorithmic Speech</i>	1360
	B. <i>Under What Circumstances Would Interferences with Algorithmic Speech be Justified?</i>	1363
	1. Prescribed by Law	1364
	2. Pursuit of a Legitimate Aim	1365
	3. Necessary in Pursuit of that Aim.....	1367
	a. Other Rights and Interests at Play	1369
	b. Issues Going Forward	1370
IV.	CONCLUSION	1373

I. INTRODUCTION

A. *Background*

Algorithms have become ubiquitous in our modern, technology-driven society. They are used in Global Position Systems (GPS), as well as in many different aspects of mobile phones and personal computers. Algorithms also assist planes in flying and cars in driving—particularly those of the self-driving variety. Despite the fact that algorithms have become a part of daily life in many ways, their operation is usually behind the scenes, and their usage goes unnoticed.

An increasing number of algorithms work to produce outputs that may be considered speech, such as automatically generated news stories, search results and their autocomplete function, as well as chat bots, such as Amazon's Alexa, Apple's Siri, Google's Assistant, and Microsoft's Cortana. There is also an untold number of bots operating on Twitter,¹ some of which Twitter has begun to prune more aggressively because of disinformation campaigns.² However, few are as infamous as Microsoft's Tay Artificial Intelligence, which was designed to mimic the speech patterns of a 19-year-old American girl.³ Within a day of its release, it was taught by users to make racist tweets; in this short time, the bot went from saying "Humans are super cool!" to "Hitler was right."⁴ These outputs were obviously not intended by the programmers.

The issues surrounding such algorithmically generated speech will only increase in importance as algorithms are developed to create more "intelligent" and complex speech,⁵ which may include unforeseen utterances. While we may not have quite reached the age where it is necessary to question whether robots should be afforded rights, we have arrived at the time when it is necessary to examine the extent to which the developers or controllers of algorithms that produce speech are protected by the right to freedom of expression.⁶

This Article aims to provide an analysis of algorithmic speech within the context of the international framework for freedom of

1. Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer & Alessandro Flammini, *Online Human-Bot Interactions: Detection, Estimation, and Characterization*, INT'L CONF. WEBLOGS & SOC. MEDIA (Mar. 27, 2017), <https://arxiv.org/abs/1703.03107> [<https://perma.cc/2Q3G-9G79>] (archived Aug. 18, 2020) (estimating that between 9% and 15% of active Twitter accounts are actually bots).

2. Andy Greenberg, *Twitter Still Can't Keep Up With Its Flood of Junk Accounts, Study Finds*, WIRED (Feb. 8, 2019), <https://www.wired.com/story/twitter-abusive-apps-machine-learning/> [<https://perma.cc/3RJ3-C83T>] (archived Aug. 20, 2020); Craig Timberg & Elizabeth Dwoskin, *Twitter is Sweeping Out Fake Accounts Like Never Before, Putting User Growth at Risk*, WASH. POST (July 7, 2018), <https://www.washingtonpost.com/technology/2018/07/06/twitter-is-sweeping-out-fake-accounts-like-never-before-putting-user-growth-risk/> [<https://perma.cc/MA7S-KRVW>] (archived Aug. 20, 2020).

3. Davey Alba, *It's Your Fault Microsoft's Teen AI Turned Into Such a Jerk*, WIRED (Mar. 25, 2016), <https://www.wired.com/2016/03/fault-microsofts-teen-ai-turned-jerk/> [<https://perma.cc/EZ33-CLLB>] (archived Aug. 20, 2020).

4. John West, *Microsoft's Disastrous Tay Experiment Shows the Hidden Dangers of AI*, QUARTZ (Apr. 2, 2016), <https://qz.com/653084/microsofts-disastrous-tay-experiment-shows-the-hidden-dangers-of-ai/> [<https://perma.cc/F3JC-EHV4>] (archived Aug. 20, 2020).

5. Steps have been taken in this direction. See Ronald Ashri, *Just how big a deal is Google's New Meena Chatbot Model?*, VENTUREBEAT (Feb. 1, 2020), <https://venturebeat.com/2020/02/01/just-how-big-a-deal-is-googles-new-meena-chatbot-model/> [<https://perma.cc/HXB8-29FF>] (archived Aug. 20, 2020).

6. Throughout this article, I will often refer to the controllers of algorithms. This is relevant because an algorithm may not always be used by only its developer; it may in fact be licensed to other parties for use.

expression. Previous literature has largely focused on certain forms of algorithmic speech, particularly search engine results and a search engine's autocomplete function. The former has been the subject of multiple cases in the United States, and thus the focus has primarily been on where such speech lies within the First Amendment doctrine—and thus the extent to which it is protected by the Constitution. The latter has been scrutinized by various national courts across Europe. Thus, there is an apparent gap in having a more comprehensive international approach to algorithmic speech, and hence the primary research question this Article addresses is the extent to which algorithmic speech is protected under international standards of freedom of expression.

Further issues arise as well: whether algorithmically generated content should be considered speech, whether the controllers of algorithms are content providers or intermediaries, when might liability be imposed for infringing algorithmic speech, the extent to which algorithmically generated content is afforded freedom of expression protection, under what circumstances would interferences be justified, and the implications of having the freedom of expression framework apply to algorithmically generated speech.

As this Article aims to address all of these issues within the current international framework for freedom of expression, international legislation and case law—particularly from the European and Inter-American systems—will be referenced where relevant. National case law and legislation will also be examined for purposes of comparisons and distinctions, and to provide further guidance as many of these issues have yet to be examined by international courts. Academic literature, as well as practical and sociological aspects relating to algorithmic speech, will be analyzed and incorporated in various areas. Recommendations will be made where it is apparent that the framework is ill-equipped to adequately deal with these issues.

It should be noted that there are a number of ways that algorithms interact with freedom of expression, which abut the topic presented in this Article, that may also be cause for concern. For instance, the use of algorithms in how news and information is presented to users may have an impact on the right to receive information, in that they can result in “echo chambers” or “filter bubbles.”⁷ While aspects such as these are no doubt worthy of investigation, they are outside the scope of this Article.

After defining algorithmic speech and introducing the variants that will form the basis of this Article, Part II will discuss algorithmic

7. COUNCIL OF EUROPE COMM. OF EXPERTS ON INTERNET INTERMEDIARIES, ALGORITHMS AND HUMAN RIGHTS: STUDY ON THE HUMAN RIGHTS DIMENSIONS OF AUTOMATED DATA PROCESSING TECHNIQUES AND POSSIBLE REGULATORY IMPLICATIONS 17 (2018).

speech and the scope of internationally recognized freedom of expression standards, as well as where algorithmic speech fits within this framework. Part III will analyze the extent to which algorithmic speech is worthy of protection under these standards.

B. What is Algorithmic Speech?

As seen above, algorithms can perform a multitude of functions in a wide range of industries and have been defined in a variety of ways over time.⁸ In this Article, the usage of the term “algorithm” will be “a set of instructions designed to produce an output.”⁹ Further, as algorithms may exist outside of the computer-centric world we live in today, usage will only encompass the common understanding of the term, in that it will refer to the algorithms that are implemented by computers.¹⁰

One may assume that if the definition of an algorithm is unsettled,¹¹ then there is likewise no single accepted definition of what constitutes algorithmic speech. Indeed, this is a vague and imprecise categorization.

In some instances, the speech or expression of algorithms is quite apparent, especially when it mimics what a person would do. This is the case with chat bots such as those that provide technical support or Microsoft’s Zo (the successor to Tay),¹² or algorithms that are fed data in order to piece together news stories.

At the opposite end of the spectrum are algorithms that are clearly not speech, such as those that perform operations in programs with no visible output. An example of this would be the algorithms on a mobile phone that determine which Wi-Fi access point to connect to when there are multiple available.¹³

8. *Algorithm Characterizations*, WIKIPEDIA, https://en.wikipedia.org/wiki/Algorithm_characterizations (last visited May 10, 2019) [<https://perma.cc/Y4EC-X6HE>] (archived Aug. 20, 2020) [hereinafter *Algorithm Characterizations*].

9. Stuart M. Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1447 n.4 (2013). Among the many definitions that I have read, Benjamin’s is among the most concise and easy to understand, particularly for those who may not have a good understanding of technical subjects such as these. *See id.*

10. *See id.*

11. *See Algorithm Characterizations*, *supra* note 8.

12. *Let’s Talk about Zo*, MICROSOFT, <https://www.zo.ai/> (last visited Feb. 29, 2019) [<https://perma.cc/K5KN-H7EY>] (archived Aug. 20, 2020).

13. *See* Dongsu Han, David G. Andersen, Michael Kaminsky, Konstantina Papagiannaki & Srinivaan Seshan, *Access Point Localization Using Local Signal Strength Gradient*, in *PASSIVE AND ACTIVE NETWORK MEASUREMENT*, 5448 LNCS 99 (2009); *see also* Kirn Gill, *Does Your Phone Use Algorithms to Decide Which Cell Tower It Should Connect To?*, QUORA (Jan. 15, 2017), <https://www.quora.com/Does-your-phone-use-algorithms-to-decide-which-cell-tower-it-should-connect-to> [<https://perma.cc/8A5S-FG76>] (archived Aug. 20, 2020).

Lying somewhere in between these two extremes are algorithms that could feasibly be considered speech, such as a search engine autocomplete function¹⁴ or the search engine results themselves. The former has been the subject of court cases in Europe,¹⁵ and the latter has been the subject of court cases and academic debate in the United States.¹⁶

Regarding the autocomplete function, courts in France have held that it does not constitute speech. In one case, a narrow interpretation of the Convention was used to find that freedom of expression is a right that only applies to “persons,” and thus it cannot be invoked in order to protect the output of an algorithm.¹⁷ In another case, it was found that an autocomplete function’s word associations are only a technical method to facilitate a search and are not expressions of opinion.¹⁸ However, the German Federal Court of Justice—the court of last resort—found that word associations, such as those resulting from an autocomplete suggestion, impart meaning.¹⁹

In the United States, courts have generally held that search engine results constitute speech,²⁰ even though search engine results merely present content provided by others. Academics have argued that algorithms are speech in that “algorithms themselves inherently incorporate the search engine company engineers’ judgments about what material users are most likely to find responsive to their queries.”²¹ Others have contended that this algorithmic output does not constitute speech due to it containing a low degree of

14. The autocomplete function I’m referring to here is utilized on Google’s search engine, among others. Once you start typing in a search string, the search engine will present a list of predictions or suggestions so as to complete what you are searching for to save you time and/or to give you new ideas.

15. For further analysis on this point, *see* Part II.B.2.a.

16. For further analysis on this point, *see* Part II.A.2.

17. *See* M. X./Google Inc., Eric S. et Google France, Tribunal de grande instance [TGI] [ordinary court of original jurisdiction], Paris, Sept. 8, 2010 (Fr.) (decision reversed by the Court of Appeal, Dec. 14, 2011). The Court of Cassation confirmed the appeal decision on 19 February 2013. In addition to the fact that the case was overturned, it should be noted that the reading of this court is quite narrow: it ignored the fact that the right to receive information as part of freedom of expression. *See id.*

18. Cour de Cassation [Cass.] [supreme court for judicial matters], 1e civ., Feb. 19, 2013, Bull. civ. I, No. 19 (Fr.) (Pierre B. v. Google Inc.); Cour de Cassation [Cass.] [supreme court for judicial matters] 1e civ., 19 June 2013, Bull. civ. I, No. 625 (Fr.) (Google v. Lyonnaise de garantie).

19. Bundesgerichtshof [BGH] [Federal Court of Justice] May 14, 2013, VI ZR 269/12 (Ger.), <http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=64163&pos=0&anz=1> [https://perma.cc/KQW5-2SPP] (archived July 11, 2020).

20. *Zhang v. Baidu.com, Inc.*, 10 F. Supp. 3d 433, 435 (S.D.N.Y. 2014); *Langdon v. Google, Inc.*, 474 F. Supp. 2d 622, 629–30 (D. Del. 2007); *Search King, Inc. v. Google Tech., Inc.*, No. 02-1457, 2003 WL 21464568, at *3 (W.D. Okla. May 27, 2003).

21. *See* Eugene Volokh & Donald M. Falk, *Google First Amendment Protection for Search Engine Search Results*, 8 J.L. ECON. & POL’Y 883, 884 (2012).

expressiveness,²² or because it should be classified as a communicative tool under the First Amendment's functionality doctrine.²³

Regardless of the arguments made on both sides of the debate,²⁴ for present purposes, this Article presumes that a search engine's autocomplete function as well as a search engine's results are forms of speech.

Algorithmic speech can take a number of different forms. The categories suggested below are by no means exclusionary, and there is certainly overlap between them—they may be more properly conceptualized as a sliding scale. However, having a conceptual understanding may aid in analyzing the issues at hand.

Form of Algorithmic Speech	Example(s)
<i>Curated production</i> —these are fed data internally	• <i>News stories</i>—more commonly used in sports news, but expanding to other areas as well, these algorithms are fed facts in order to produce stories that read as though they were written by a human²⁵ • <i>Search engine results</i>—using predefined criteria, search engines use algorithms (and many times combinations of them) in order to display the most relevant results in the provider's estimation in response to an external source of a string of text provided by the user
<i>Interactive/responsive production</i> —these respond to data from external sources	• <i>Chat bots</i>—many chat programs, whether in social media messaging or customer support, utilize algorithms to respond to people, often with the intent to imitate a person; Microsoft's Tay could be considered an example of this, but could also fall into the following category

22. See generally Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149 (2008).

23. See generally Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495 (2013).

24. See *infra* notes 43, 44, 45, 116, 118 & 120 and accompanying text.

25. Matthew Jenkin, *Written Out of the Story: The Robots Capable of Making the News*, GUARDIAN (July 22, 2016), <https://www.theguardian.com/small-business-network/2016/jul/22/written-out-of-story-robots-capable-making-the-news> [https://perma.cc/EL4E-ZXRV] (archived Aug. 20, 2020).

<i>Semi-autonomous production</i> —these also respond to data from external sources but have more “freedom” to produce unexpected results from what the programmers intended.	• Tay “learned,” or rather adapted, autonomously based upon the input of users who interacted with it • <i>Search engines’ autocomplete functions</i>—these incorporate the input of many people who searched for certain strings of text without direct oversight from the programmers of the algorithm
<i>Fully autonomous production</i> —the scenario in which an algorithm produces speech fully independent of human intervention or input ²⁶	• Not currently in existence

As the last of these categories does not currently exist—and is unlikely to exist for some time—this Article will focus on the algorithms that would fall within the first three categories above: curated production, interactive/responsive production, and semi-autonomous production.²⁷ This list also does not purport to contain all forms of algorithmic speech; it merely exemplifies some of the more well-known forms, around which the discussion will develop. Several of the specific examples of algorithmic speech given above will be examined in more detail below within the context of the international human rights framework.

26. While this may seem far-fetched, it may not be so far off as once thought. See Adrienne LaFrance, *An Artificial Intelligence Developed Its Own Non-Human Language*, ATLANTIC (June 15, 2017), <https://www.theatlantic.com/technology/archive/2017/06/artificial-intelligence-develops-its-own-non-human-language/530436/> [https://perma.cc/SB8T-RDG7] (archived Aug. 20, 2020); Timothy Revell, *Google’s Neural Networks Invent Their Own Encryption*, NEW SCIENTIST (Oct. 26, 2016), <https://www.newscientist.com/article/2110522-googles-neural-networks-invent-their-own-encryption/> [https://perma.cc/6HTE-XTTM] (archived Aug. 20, 2020).

27. For reasons unrelated to Skynet, we are unlikely to have fully autonomous AI “out in the wild” in the near future. Ethical and safety standards need to be developed, and an optimistic prediction for human-level artificial intelligence is the year 2029. See Ray Kurzweil, *Don’t Fear Artificial Intelligence*, TIME (Dec. 19, 2014), <http://time.com/3641921/dont-fear-artificial-intelligence/> [https://perma.cc/V4FQ-DDGM] (archived Aug. 20, 2020).

II. ALGORITHMIC SPEECH AND THE SUBSTANTIVE SCOPE OF FREEDOM OF EXPRESSION

This Article focuses on algorithmic speech within the context of the international framework for freedom of expression. After briefly introducing this framework, court cases and arguments made by academics within the markedly different framework of the United States will be examined to provide further context, before returning to evaluate how different forms of algorithmic speech fit within the international framework, in regard to attribution and their classification as content providers or intermediaries, and liability for harmful speech.

A. Frameworks for Freedom of Expression

1. The International Framework

The Universal Declaration of Human Rights (UDHR) has formed the foundation of many human rights instruments that have followed in its wake.²⁸ Freedom of opinion and expression is specifically guaranteed in this document,²⁹ and it has been further enshrined in international treaties and developed through the case law of international bodies and regional courts.

The International Covenant on Civil and Political Rights (ICCPR),³⁰ the European Convention on Human Rights (ECHR),³¹ the American Convention on Human Rights (ACHR),³² and the Charter of Fundamental Rights of the European Union³³ all provide protection for the right to freedom of expression, albeit with some limitations. This right is extremely important and has been held to be “a cornerstone of the survival of a democratic society.”³⁴ Generally, the right includes

28. *The Foundation of International Human Rights Law*, U.N., <https://www.un.org/en/sections/universal-declaration/foundation-international-human-rights-law/index.html> (last visited Feb. 29, 2020) [<https://perma.cc/4ELX-Y3EN>] (archived Aug. 20, 2020). The American Declaration of the Rights and Duties of Man also served as guiding principles for the American Convention on Human Rights. *See id.*

29. G.A. Res. 217 (III), art. 19, Universal Declaration of Human Rights (Dec. 10, 1948) [hereinafter UDHR].

30. International Covenant on Civil and Political Rights art. 19, *opened for signature* Dec. 16, 1966, 999 U.N.T.S. 171 (entered into force Mar. 23, 1976) [hereinafter ICCPR].

31. Convention for the Protection of Human Rights and Fundamental Freedoms art. 10, Nov. 4, 1950, 312 E.T.S. 5 [hereinafter ECHR].

32. American Convention on Human Rights art. 13, Nov. 22, 1969, 1144 U.N.T.S. 123 [hereinafter ACHR].

33. Charter of Fundamental Rights of the European Union art. 11, Dec. 7, 2000, 2000 O.J. (C 364) 1 [hereinafter CFREU].

34. *Usón Ramírez v. Venezuela*, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 207, ¶ 47 (Nov. 20, 2009). Here, the

the ability to “receive and impart information and ideas” through any media, “regardless of frontiers.”³⁵

This latter clause, “regardless of frontiers,” refers to the application of this standard to all speech that crosses borders. Thus, speech that is transmitted over the Internet should be given the same freedom of expression protection as domestic speech, regardless of the place of origination.³⁶ This is particularly important to algorithmic speech in the sense that many current forms of algorithmic speech originate from servers located in other countries. For instance, despite Google’s web search being so popular that its name is often referred to as a replacement for the service itself (e.g., “to google something”), Google only operates twenty-one servers around the world, more than half of which are located in the United States.³⁷

It is also important to note that freedom of expression protection applies to information and ideas that may “offend, shock or disturb,” and is not restricted only to those “that are favourably received or regarded as inoffensive or as a matter of indifference.”³⁸

Court emphasized that this is true “particularly in matters of public interest” and referred to “its jurisprudence established in numerous cases.” *See id.*

35. This language is found in all of the aforementioned documents. *See supra* notes 30–33.

36. ICCPR, *supra* note 30, art. 19(2); CFREU, *supra* note 33, art. 11(1); ECHR, *supra* note 31, art. 10(1). *See also* JAN OSTER, MEDIA FREEDOM AS A FUNDAMENTAL RIGHT 60–70 (2015) [hereinafter OSTER, MEDIA FREEDOM]; JAN OSTER, EUROPEAN AND INTERNATIONAL MEDIA LAW 39 (2017) [hereinafter OSTER, MEDIA LAW]. Cf. *Cox v. Turkey*, Eur. Ct. H.R. App. No. 2933/03, ¶ 31 (2010), where an American citizen was denied re-entry into Turkey for comments made about the Armenian genocide; the Court stated “that the ban on the applicant’s re-entry is materially related to her right to freedom of expression because it disregards the fact that Article 10 rights are enshrined ‘regardless of frontiers’ and that no distinction can be drawn between the protected freedom of expression of nationals and that of foreigners.” *Id.*

37. Google Staff, *Data Centers*, GOOGLE, <https://www.google.com/about/datacenters/inside/locations/index.html> (last visited Feb. 22, 2020) [<https://perma.cc/AY4P-TWQA>] (archived Aug. 20, 2020). This number has increased fairly dramatically over time—as of May 24, 2019, Google only had 16 data centers. *See id.*

38. *Handyside v. United Kingdom*, Eur. Ct. H.R. App. No. 5493/72, ¶ 49 (1976). This principle is reiterated in, amongst others, *Sunday Times v. United Kingdom* (No. 1), Eur. Ct. H.R. App. No. 6538/74, ¶ 65 (1979); *Lingens v. Austria*, Eur. Ct. H.R. App. No. 9815/82, ¶ 41 (1986); *Thorgeir Thorgeirson v. Iceland*, Eur. Ct. H.R. App. No. 13778/88, ¶ 63 (1992); *Axel Springer AG v. Germany* (No. 1), Eur. Ct. H.R. App. No. 39954/08, ¶ 78 (2012). This quotation was also used by the Inter-American Court in the case of “The Last Temptation of Christ” (*Olmedo-Bustos v. Chile*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 73, ¶ 47 (Feb. 5, 2001)), which was later referenced in a number of cases. *See, e.g.,* *Kimel v. Argentina*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 177, ¶ 88 (May 2, 2008); *Canese v. Paraguay*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 111, ¶ 83 (Aug. 31, 2004); *Herrera-Ulloa v. Costa Rica*, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 107, ¶ 113 (July 2, 2004); *Ivcher-Bronstein v. Peru*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 74, ¶ 152 (Feb. 6, 2001).

However, the right to freedom of expression is not completely unconstrained. Common limitations across these instruments are for “national security, public order, or public health or morals,” and the right may not be fully realized if it comes into direct conflict with the rights of another person.³⁹ Therefore, the analysis of a supposed freedom of expression infringement focuses on whether the interference was justified, taking into account the relevant conflicting rights and interests. This framework will be examined in further detail in Part 3.

Relatively little has been said in international jurisprudence about freedom of expression on the Internet, much less algorithmic speech. However, certain functions, such as the maintenance of Internet news archives, has been explicitly held to be covered by Article 10 of the ECHR:

The Court has consistently emphasised that Article 10 guarantees not only the right to impart information but also the right of the public to receive it. In light of its accessibility and its capacity to store and communicate vast amounts of information, the Internet plays an important role in enhancing the public’s access to news and facilitating the dissemination of information generally. The maintenance of Internet archives is a critical aspect of this role and the Court therefore considers that such archives fall within the ambit of the protection afforded by Article 10.⁴⁰

One could argue that similar logic could be extrapolated to cover the algorithms used in search engines, for instance. Going forward, this Article will examine how algorithmic speech, currently in its infancy, has been viewed by academics and courts in the United States—as academics have written on the issue fairly extensively and there are a number of cases concerning search engines—before returning to the international framework.

2. The United States’ Framework and Algorithmic Speech

The United States examines the right to freedom of expression (or rather freedom of speech) in quite a different manner than that just described. While the US approach looks, *a priori*, at a particular act to determine whether it qualifies as speech and is thus entitled to protection, the international approach, as stated above, looks at interferences to speech and whether they can be justified when taking into account the relevant rights and interests. Regardless of the framework, however, it may be useful to look at how commentators and

39. The quoted language is taken directly from the ACHR, although the ICCPR uses almost identical wording, and the ECHR’s language is very similar and touches upon the same exceptions. See ACHR, *supra* note 32, art. 12, ¶ 3; see also ICCPR, *supra* note 30, art. 19 ¶ 3(b); ECHR, *supra* note 31, art. 10 ¶ 2.

40. *Times Newspapers, Ltd. v. United Kingdom* (Nos. 1 & 2), Eur. Ct. H.R. App. Nos. 3002/03 & 23676/03, ¶ 27 (2009) (citations omitted).

courts have approached algorithmic speech in the United States. In several instances, federal district courts have held that algorithmic speech is speech and thus entitled to protection.

In the United States, freedom of speech is a right enshrined in the First Amendment to the Constitution,⁴¹ and it is typically broader than the right to freedom of expression found internationally. Courts look to whether an act can be considered “speech” and thus whether it should be afforded protection under the First Amendment. There are limitations, which are categorical in nature, and they are relatively narrow in comparison to the justification analysis and balancing of rights and interests utilized internationally.⁴²

There have been quite a number of articles written on the extent to which algorithmic speech is protected by the First Amendment in the United States, usually within the context of search engine results. This has resulted in a vigorous debate with proponents on all areas of the spectrum, advocating for a variety of theories with which to approach the issue. Several academics have argued that the algorithmic speech of search engines is protected by the First Amendment.⁴³ On the other hand, others have contended that this algorithmic output should not be protected.⁴⁴ Still others argue that a

41. U.S. CONST. amend. I. The amendment states: “Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.” *See id.*

42. The extent to which speech may be prohibited or merely limited differ between the categories, which include: obscenity (*Miller v. California*, 413 U.S. 15 (1973)); fighting words and offensive speech (*Chaplinsky v. New Hampshire*, 315 U.S. 568 (1942)); false statements of fact (*Gertz v. Robert Welch, Inc.*, 418 U.S. 323 (1974)); child pornography (*New York v. Ferber*, 458 U.S. 747 (1982)); speech that incites imminent lawless action (*Brandenburg v. Ohio*, 395 U.S. 444 (1969)); speech owned by others such as through copyright or trademarks (*Harper & Row, Publishers v. Nation Enters.*, 471 U.S. 539 (1985)); and commercial speech such as advertising (*Cent. Hudson Gas & Elec. Corp. v. Pub. Serv. Comm’n*, 447 U.S. 557 (1980)). Additionally, courts presume any restriction on speech to be invalid and the onus is on the government to convince the court that the restriction is constitutional. For a look at how this is examined internationally, see *infra* Part III.B.3 (discussing the “necessary in pursuit of the aim” justification analysis) and Part III.B.3.a (describing the balancing of rights at play within judicial oversight of freedom of expression, particularly for cases involving algorithmic speech).

43. *See generally* Benjamin, *supra* note 9 (maintaining that current First Amendment jurisprudence should be understood to cover a broad spectrum of algorithmic output, especially those that involve a substantive communication); Volokh & Falk, *supra* note 21 (contending that search engines exercise editorial judgment in determining what information to convey to the user, and that they are analogous to newspapers and book publishers and therefore protected by the First Amendment).

44. *See generally* Bracha & Pasquale, *supra* note 22 (observing that speech with a low degree of expressiveness is commonly excluded from First Amendment protection and that search engine results are less expressive than these categories of speech that are excluded, in addition to the fact that these results are a form of speech that do not realize First Amendment values despite them having a communicative function); Wu,

more graduated or nuanced approach should be utilized, where algorithmic speech should be protected in certain instances and denied that protection in others.⁴⁵

In court, Google has repeatedly argued that its search results are protected speech and thus protected by the First Amendment. In 2003, Google argued in *Search King, Inc. v. Google Technology, Inc.* that its PageRank results were subjective opinions.⁴⁶ Search King offered search optimization to clients, and when Google discovered this, it demoted the clients' ranking in its search results. In turn, Search King sued Google for tortious interference with contract. The court found Google's argument persuasive and held that Google's PageRanks did not "contain provably false connotations" and were therefore opinions entitled to "full constitutional protection."⁴⁷ In another instance, a different court found that an injunction sought that would shape Google's search results would violate its First Amendment rights.⁴⁸

Another case involved the largest search engine provider in China—Baidu. At the request of the Chinese government, Baidu blocked results concerning the prodemocracy movement in China from appearing in search results in the United States.⁴⁹ Whereas the previous two cases engaged in little analysis on this issue, the court in *Zhang v. Baidu.com Inc.* more thoroughly discussed this topic,⁵⁰ and found that "there is a strong argument to be made that the First Amendment fully immunizes search-engine results from most, if not

supra note 23 (arguing that the First Amendment's functionality doctrine precludes coverage from carriers/conduits and communicative tools; as such, search engines should typically be classified as a tool as opposed to speech, and automated concierge services as well, unless the opinions of the programmer are reflected in the output).

45. See generally Michael J. Ballanco, Comment, *Searching for the First Amendment: An Inquisitive Free Speech Approach to Search Engine Rankings*, 24 GEO. MASON U. C.R.L.J. 89 (2013) (advancing a fact-based analysis of whether the search engine presents relatively neutral results, and if it is found that the search engine is advancing its own commercial interest it should be considered commercial speech and hence entitled to less protection by the First Amendment); Josh Blackman, *What Happens if Data Is Speech?*, 16 U. PA. J. CONST. L. ONLINE 25 (2014) (proposing a framework that focuses on the nexus between algorithmic outputs and human interaction; with more human interaction the output will be closer to what the human created herself and thus deserving of protection, whereas if the output is relatively autonomous with little human involvement it lies farther away from human expression that warrants protection); James Grimmelman, *Speech Engines*, 98 MINN. L. REV. 868 (2014) (positing that a search engine is neither a conduit that is categorically not entitled to First Amendment protection or an editor that is, but an advisor that should not receive protection where it deceives the user that it is supposed to inform).

46. *Search King, Inc. v. Google Tech., Inc.*, No. 02-1457, 2003 WL 21464568, at *3 (W.D. Okla. May 27, 2003).

47. *Id.* at *4.

48. *Langdon v. Google, Inc.*, 474 F. Supp.2d 622, 629–30 (D. Del. 2007).

49. *Zhang v. Baidu.com, Inc.*, 10 F. Supp.3d 433, 435 (S.D.N.Y. 2014), *appeal withdrawn* (2d Cir. 2014).

50. *Id.* at 436–43.

all, kinds of civil liability and government regulation.”⁵¹ The court went on to discuss an argument for examining search engine results under merely an intermediate level of scrutiny.⁵² Ultimately, the court did not decide exactly which level of protection search engine results should be afforded generally, but found that the intermediate scrutiny test was inapplicable to the current case.⁵³

Google has made similar arguments in more recent cases. In a 2017 case, Google had delisted a number of e-ventures’ websites from its search results for violating its guidelines; Google was granted summary judgment on the grounds that formulating search results—including deciding which links to list and how to order or rank them—are essentially editorial decisions protected by the First Amendment.⁵⁴ In 2019, in a case where a stock image company sued Google because of its displeasure with how its ranking in search results had fallen precipitously several years prior, Google moved for a judgment on the pleadings partially upon the aforementioned arguments.⁵⁵ Noting that no appellate court had examined this issue, the court found that even if search engines were generally protected, Google “[could not] hide behind the First Amendment”—breach of contract could still occur, and discovery would illuminate what in fact happened.⁵⁶

Amazon has also made similar arguments in a legal memorandum submitted for a criminal case.⁵⁷ Here, police attempted to obtain a search warrant to procure the voice recording, taken by Amazon through its Alexa service, of the prime suspect in a murder

51. *Id.* at 438. The court outlined the principles it used as such: “First, as a general matter, the Government may not interfere with the editorial judgments of private speakers on issues of public concern—that is, it may not tell a private speaker what to include or not to include in speech about matters of public concern. Second, that rule is not ‘restricted to the press, being enjoyed by business corporations generally and by ordinary people engaged in unsophisticated expression as well as by professional publishers.’ Third, the First Amendment’s protections apply whether or not a speaker articulates, or even has, a coherent or precise message, and whether or not the speaker generated the underlying content in the first place. And finally, it does not matter if the Government’s intentions are noble—for example, to promote ‘press responsibility,’ or to prevent expression that is ‘misguided, or even hurtful.’” *Id.* at 437–38 (citations omitted).

52. *Id.* at 439–41. The argument was originally made in Bracha & Pasquale, *supra* note 22, at 1191–94, which relied upon the intermediary scrutiny used by the Supreme Court when examining regulations of cable television operators that required the operators to carry local broadcast stations in *Turner Broad. Sys., Inc. v. FCC*, 512 U.S. 622, 662 (1994). Typically, content-based speech restrictions are reviewed under strict scrutiny, and content-neutral restrictions under intermediate scrutiny.

53. *Zhang*, 10 F. Supp. 3d at 439–41.

54. *e-ventures Worldwide, LLC v. Google, Inc.*, No. 2:14-cv-646-FtM-PAM-CM, 2017 WL 2210029, at *8–9 (M.D. Fla. Feb. 8, 2017).

55. *Dreamtime.com, LLC v. Google, LLC*, No. C 18-01910 WHA, 2019 WL 2372280, at *1–2 (N.D. Cal. June 5, 2019).

56. *Id.* at *3–4.

57. Memorandum of Law in Support of Amazon’s Motion to Quash Search Warrant, *Arkansas v. Bates*, No. CR-2016-370-2 (Benton Cty. Cir. Ct., Feb. 17, 2017) [hereinafter Amazon’s Memorandum].

investigation.⁵⁸ Amazon argued that both the speech submitted to Alexa by the user, as well as the responses generated by Alexa,⁵⁹ are protected by the First Amendment and thus subject to heightened scrutiny by a court.⁶⁰ In the end, the court did not have to rule on the matter as the defendant agreed to release the recordings.⁶¹

Despite the categorical approach to freedom of expression in the United States, the foregoing discussion shows that academics and courts have struggled with analyzing algorithmic outputs, and the struggle will continue as new forms emerge and claims for protection are made. While current case law points in the direction that the algorithmic output of search engines is constitutionally-protected speech, the law is far from settled. The lack of clarity on this issue equally applies—and perhaps even more so—to the international framework, which we will return to in the following section.

B. How Might Algorithmic Speech Fit into the International Framework

1. Attribution of Algorithmic Speech and Status as a Content Provider or Intermediary

Another crucial question that must be answered is whether the creators or controllers of the programs that produce algorithmic speech should be considered content providers or intermediaries. This

58. *Id.* See also Debra Cassens Weiss, *Alexa's Responses to Customers Are Protected by The First Amendment, Amazon Argues in Murder Case*, ABA J. (Feb. 27, 2017, 7:00 AM CST), http://www.abajournal.com/news/article/alexa_responses_to_customers_are_protected_by_the_first_amendment_amazon_a/ [<https://perma.cc/J6DF-8UV8>] (archived Aug. 20, 2020).

Alexa is an interactive cloud service where users talk to an Alexa-enabled device in order to “play music, answer general questions, set an alarm or timer and more.” *Alexa*, AMAZON DEVELOPER (Sept. 21, 2017), <https://web.archive.org/web/20170921015141/https://developer.amazon.com/alexa> (last visited Sept. 28, 2020) [<https://perma.cc/TZK8-6PDF?type=image>] (archived Sept. 28, 2020).

59. For this latter argument, Amazon cited a couple of the aforementioned cases such as *Search King* and *Baidu*. See Amazon’s Memorandum, *supra* note 57, at 11–12.

60. Amazon argued that due to the heightened scrutiny, “it is the government’s burden to show both that (1) it has a ‘compelling interest’ in the requested information and (2) there is a ‘sufficient nexus’ between the information sought and the underlying inquiry of the investigation.” *Id.* at 12.

61. Allison Grande, *Amazon Turns Over Recordings With Murder Suspect’s OK*, LAW360 (Mar. 7, 2017, 8:11 PM EST), <https://www.law360.com/articles/899149/amazon-turns-over-recordings-with-murder-suspect-s-ok> [<https://perma.cc/5HZZ-GPKB>] (archived Aug. 20, 2020); Press Release, Kathleen T. Zellner & Associates, P.C., Amazon Echo Subpoena in Arkansas Murder Case (Mar. 6, 2017), <https://arstechnica.co.uk/wp-content/uploads/2017/03/echoagreement.pdf> [<https://perma.cc/6F7Q-V8NT>] (archived Aug. 20, 2020).

determination ultimately has implications for liability, which will be examined in the next subsection.

The authors or creators of information are considered content providers, which may include publishers, news outlets, bloggers, or even creators of YouTube videos.⁶² On the other hand are mere speech intermediaries or transmitters, such as communication networks, newspaper vendors, search engines, social networks, and news aggregators.⁶³ There is therefore a distinction between a content provider or “the media” and a “medium”; the primary differentiator between the two lies in the former’s exercise of editorial control, which is the “creation, selection or redaction of content before its publication.”⁶⁴ Furthermore, persons or entities are not to be considered mere intermediaries if they provide their own content, adopt third-party content, or initiate the dissemination or publication of third-party content.⁶⁵

As algorithmic speech comes in many shapes and forms, it is not immediately clear whether the creator of the speech should be categorized as a content provider or a mere intermediary. At first glance, it may seem clear that algorithmic speech is attributable merely to the person—or company that employs the person—who programmed the algorithm, or to the entity in control of the algorithm, which in turn would deem that person or company the author. However, this may not necessarily hold true in all instances.

Certain forms are relatively straightforward. For instance, the publishers of automatically generated news stories no doubt exercise editorial control over the content and would therefore be considered content providers, regardless of whether they created the algorithm originally. Similarly, with basic chat bots, where an algorithm responds to user-submitted text with scripts prepared by either the creator or controller—that entity is thus providing the content—hence

62. It is important to note the ‘Internet content provider’ may have a slightly different understanding in common parlance. In the EU, a ‘content provider’ is “the information source under communication theory.” Jan Oster, *Communication, Defamation and Liability of Intermediaries*, 35 *LEGAL STUD.* 348, 351 (2015) [hereinafter Oster, *Liability of Intermediaries*]. In the U.S., ‘information content provider’ is defined as “any person or entity that is responsible, in whole or in part, for the creation or development of information provided through the Internet or any other interactive computer service.” Communications Decency Act of 1996, Pub. L. No. 104-104, sec. 509, § 230(e)(3), 110 Stat. 56, 139 (1996).

63. OSTER, *MEDIA FREEDOM*, *supra* note 36, at 57.

64. *Id.* at 58; *see also* Directive 2010/13/EU, of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) Recitals 25 and 26; Directive 2002/21/EC, of the European Parliament and of the Council of 7 March 2002 on a common regulatory framework for electronic communications networks and services (Framework Directive) art. 2(c).

65. OSTER, *MEDIA LAW*, *supra* note 36, at 14; Oster, *Liability of Intermediaries*, *supra* note 62, at 358.

they should be deemed content providers as there is editorial control in the text that is ultimately presented to the user.

Other forms of algorithmic speech are not so straightforward. Attribution may become a bit more complicated when looking at adaptive algorithms (semi-autonomous production), such as more advanced chat bots, Microsoft's Tay, or a search engine's autocomplete function, where the algorithmic speech production is a compilation of the instant interaction combined with many interactions that had occurred previously.⁶⁶ Thus when Tay started making racist comments less than 24 hours after it was launched,⁶⁷ it was not only the result of the programmers' algorithm but also of all those people who interacted with it. Hence Microsoft is arguably not providing its own content—at least not in whole. The company undoubtedly did not intend for Tay to make comments such as “Hitler was right I hate the jews” and disabled the Twitter account after only a day of being “in the wild.”⁶⁸ On the other hand, several bad actors did intend to “game” Tay so as to make it speak the way it did. Microsoft *did* program Tay, but in a scenario such as this, should Tay's speech be solely attributable to Microsoft?

One could argue that because users interacted with Tay's algorithm in an abusive manner, the speech should not be attributable to Microsoft. However, if someone is injured by hate speech or defamation, etc., the question would remain as to who should be held liable when the output is the amalgamation of many different users' input, who may oftentimes be anonymous.⁶⁹

This scenario—which would also apply to a search engine's autocomplete function—would require analysis under whether one's own content was provided or whether third-party content was adopted.⁷⁰ This is an objective standard based upon the perception of an ordinary reasonable person.⁷¹ A third-party statement being adopted may be indicated by “whether the publisher invited the statement, expressly approved of them or attached his brand name to

66. Search Using Autocomplete, GOOGLE, <https://support.google.com/websearch/answer/106230> (last visited Feb. 29, 2020) [<https://perma.cc/UEN7-WJEC>] (archived Aug. 20, 2020) (stating that autocomplete takes into account the text string that was entered, a user's relevant past searches, and what other users are searching for, including trending stories). Interestingly, Google claims that these suggestions “[a]re not statements by other people or Google about [one's] search terms.” *Id.*

67. Helena Horton, *Microsoft Deletes 'Teen Girl' AI After It Became a Hitler-Loving Sex Robot Within 24 Hours*, TELEGRAPH (Mar. 24, 2016, 3:37 PM), <http://www.telegraph.co.uk/technology/2016/03/24/microsofts-teen-girl-ai-turns-into-a-hitler-loving-sex-robot-wit/> [<https://perma.cc/L8NM-U6Z8>] (archived Aug. 20, 2020).

68. *Id.*

69. For a more thorough discussion on these issues, see *infra* Parts II.B.2., concerning liability, and III.B.3.a., concerning other rights and interest at play in the balancing exercise.

70. Oster, *Liability of Intermediaries*, *supra* note 62, at 359.

71. *Id.* at 358.

them.”⁷² Consequently, neither providing your own content nor adopting a third party’s requires that a particular statement is endorsed, and persons or entities could be considered content providers even if the statement is not reflective of their opinion.⁷³ To an ordinary reasonable person, it likely appears that Microsoft Tay or the search engine’s autocomplete function are presenting new content (or are at least adopting third-party content), and thus they should be considered content providers even if they do not officially support the output.⁷⁴ However, the fact that this standard is a bit of a moving target may change the analysis—people may become more tech-savvy and informed, which would make an ordinary reasonable person realize that this algorithmic output is not content provided by Microsoft Tay or the search engine. Search engines could also potentially circumvent this by showing a large notice informing users, when they are searching, that the suggestions are merely trending text strings of other users.

Even where users are not purposefully attempting to game the algorithm or interact with it in an abusive manner, it may result in a breach of the law. The output of a search engine’s autocomplete function has been found to be defamatory,⁷⁵ as have web and image search results,⁷⁶ and image search results appear to be discriminatory in some instances.⁷⁷ In some sense, it seems unjust to consider these

72. *Id.* at 359. *See also* Law Soc’y v. Kordowski [2011] EWHC (QB) 3185 (Eng.).

73. Oster, *Liability of Intermediaries*, *supra* note 62, at 359.

74. *See* BGH, May 14, 2013, VI ZR 269/12, 9 (Ger.), <http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=64163&pos=0&anz=1> [<https://perma.cc/KQW5-2SPP>] (archived July 11, 2020) (finding that users expect that “the search queries completed through the suggested word combination reflect content-related relationships”); Karl-Nikolaus Peifer, *Google’s Autocomplete Function—is Google a Publisher or a Mere Technical Distributor?*, 3 QUEEN MARY J. INTELL. PROP. 318 (2013). However, this principle this decision is not accepted in all jurisdictions, and there have been a number of cases in different jurisdictions that have reached divergent conclusions using a variety of reasoning. *See* Stavroula Karapapa & Maurizio Borghi, *Search Engine Liability for Autocomplete Suggestions: Personality, Privacy and the Power of the Algorithm*, 23 INT’L J.L. INFO. TECH. 261, 275–81 (2015) (discussing a number of autocomplete cases throughout Europe).

75. *See* Karapapa & Borghi, *supra* note 74, at 278–81.

76. *See* Milorad Trkulja v Google Inc LLC [No 5] (2012) VSC 533 (Austl.).

77. Matthew Kay, Cynthia Matuszek & Sean A. Munson, *Unequal Representation and Gender Stereotypes in Image Search Results for Occupations*, ASSOC. COMPUTING MACHINERY (2015), <https://www.csee.umbc.edu/~cmat/Pubs/KayMatuszekMunsonCHI2015GenderImageSearch.pdf> [<https://perma.cc/RX2K-J9JG>] (archived Aug. 20, 2020). In this study conducted in the U.S., professions were searched for in Google’s image search to for gender biases in the results. One particularly notable finding was that the results when searching for “CEO”: the percentage of women in the top 100 results was 11%, whereas the actual percentage of CEOs who are women in the U.S. is 27%. It was also found that the when exposed to the skewed results, this resulted in a feedback loop that further reinforced

algorithmic speech outputs the speech of the search engine when it is in fact largely a reflection of society's suspicions, preconceived opinions, or biases.

Even if a reasonable person does not view autocomplete suggestions as the content of the search engine, or Tay's tweets as that of Microsoft, in these instances, there may also be the exercise of editorial control. Although the output is not scripted to the same extent as basic chat bots, a programmer still designed the algorithm which generates the output,⁷⁸ and the preemptive policies used could be viewed as a form of *ex ante* editorial control,⁷⁹ such as through making certain topics or combinations of text off-limits *ab initio*.⁸⁰ Despite the fact that Google states that its search engine's autocomplete "predictions are generated by an algorithm automatically without human involvement,"⁸¹ Google redacts material that is sexually explicit, hateful, violent, or dangerous,⁸² thus showing some measure of editorial control.

At the same time, search engines may be considered mere intermediaries in regard to the algorithms that determine search results in the consideration of whether one's own content was provided or whether third-party content was adopted. Unlike the autocomplete function, to an ordinary reasonable person it is likely clear that search engine results are lists of content that are neither provided nor adopted by the search engine itself, as they only present excerpts and links to content provided elsewhere. Among those discussed here, this form of algorithm (as well as those that may have similar functionality) is the one that is most likely to succeed if the creator or controller argues that

these biases. *See also* Adrienne LaFrance, *Be Careful What You Google*, ATLANTIC (Apr. 10, 2015), <https://www.theatlantic.com/technology/archive/2015/04/be-careful-what-you-google/390207/> [<https://perma.cc/LU8Q-ECZS>] (archived Aug. 20, 2020); Jennifer Langston, *Who's a CEO? Google Image Results Can Shift Gender Biases*, U. WASH. NEWS (Apr. 9, 2015), <https://www.washington.edu/news/2015/04/09/whos-a-ceo-google-image-results-can-shift-gender-biases/> [<https://perma.cc/J476-W8DT>] (archived Aug. 20, 2020). To my knowledge, search results such as these have not been contested in court, but they have been part of the push for the need for algorithmic accountability. *See infra* Part III.A. for more information on this movement.

78. *See also* Karapapa & Borghi, *supra* note 74, at 274 (stating that one of the judicial trends holds that the autocomplete function introduces "an additional source of informative content of which the search engine is solely responsible" and thus is no longer a mere intermediary).

79. Eur. Consult. Ass., *Recommendation CM/Rec(2011)7 of the Comm. of Ministers to member states on a new notion of media*, adopted Sept. 21, 2011, ¶ 32 (2011), <https://www.coe.int/en/web/media-freedom/committee-of-ministers> [<https://perma.cc/4EY4-U5MB>] (archived Aug. 20, 2020).

80. This could have implications for self-censorship. *See infra* Part III.B.3.b.

81. *How search predictions work on Google*, GOOGLE, <https://support.google.com/websearch/answer/106230> (last visited Feb. 29, 2020) [<https://perma.cc/248W-3KKR>] (archived Aug. 20, 2020).

82. *Autocomplete Policies*, GOOGLE, <https://support.google.com/websearch/answer/7368877> (last visited Feb. 29, 2020) [<https://perma.cc/7NQG-8FQV>] (archived Aug. 20, 2020).

it is a mere intermediary and thus may avail itself of the associated limitations of liability discussed in the following subpart. However, it is possible that editorial control could be found in the determination and ranking of search results.⁸³

Finally, even if there is a complete lack of oversight or redaction, as was apparently the case with Microsoft Tay, editorial control may arguably still be found. When “one-to-many” traditional media outlets, such as broadcasters or newspapers, disseminate third-party content, and have the ability to edit the content but elect not to exercise editorial control, then they must be considered primary publishers and not mere intermediaries.⁸⁴ While autocomplete functions and bots on social media platforms do not neatly fall into this category, they essentially operate as a “one-to-many” form of communication. “Many-to-many” (often online) platforms of communication do not exercise editorial control over third-party content published, so long as they are not aware of the harmful speech being published, are not able to prevent its dissemination, and do not adopt or modify the content.⁸⁵ In any case, many of these companies have shown that they are able to largely prevent the dissemination of harmful speech.⁸⁶

Creators or controllers of these algorithms may be in a bit of a quandary—the more (editorial) control they exert in order to avoid undesirable outputs, the more they open themselves to liability.⁸⁷ Not only would the process to filter out defamatory remarks be extremely complicated in some cases,⁸⁸ it may also have further implications

83. It is interesting to note that U.S. courts have made this exact finding—that search engines exercise editorial control in determining their search results, including how they are ranked. *See, e.g.,* *e-ventures Worldwide, LLC v. Google, Inc.*, No. 2:14-cv-646-FM-PAM-CM, 2017 WL 2210029, at *8–9 (M.D. Fla. Feb. 8, 2017); *Zhang v. Baidu.com, Inc.*, 10 F. Supp. 3d 433, 437 (S.D.N.Y. 2014).

84. Oster, *Liability of Intermediaries*, *supra* note 62, at 361.

85. *Id.*

86. *See* Elizabeth Schulze, *EU Says Facebook, Google and Twitter Are Getting Faster at Removing Hate Speech Online*, CNBC (Feb. 4, 2019), <https://www.cnbc.com/2019/02/04/facebook-google-and-twitter-are-getting-faster-at-removing-hate-speech-online-eu-finds-.html> [<https://perma.cc/6VZS-7W9Y>] (archived Aug. 20, 2020). Interestingly, however, the algorithms that are used to detect such speech may themselves be biased. *See* Thomas Davidson, Debasmita Bhattacharya & Ingmar Weber, *Racial Bias in Hate Speech and Abusive Language Detection Datasets*, PROCEEDINGS OF THE THIRD WORKSHOP ON ABUSIVE LANGUAGE ONLINE 25 (Aug. 1, 2019), <https://www.aclweb.org/anthology/W19-3504.pdf> [<https://perma.cc/9VTZ-UFKQ>] (archived Aug. 20, 2020). This essentially is treating platforms like governments, but without the same level of accountability to the public. Daphne Keller, *Facebook Restricts Speech by Popular Demand*, ATLANTIC (Sept. 22, 2019), <https://www.theatlantic.com/ideas/archive/2019/09/facebook-restricts-free-speech-popular-demand/598462/> [<https://perma.cc/DJ9M-YLCW>] (archived Aug. 20, 2020).

87. This may explain Google’s explanation above regarding how autocomplete suggestions are formulated.

88. For search results, this would be notably harder than for an advanced chat bot. For instance, a search engine could create an algorithm to trawl the Internet to

down the line—holding search engines potentially liable in these instances could lead to them eventually removing features such as these.⁸⁹ On the other hand, in cases where the entity is deemed to be a provider, perhaps users interacting with the algorithm in an abusive manner could be a mitigating factor in a potential award of damages.⁹⁰ The question might ultimately become whether the benefits (such as to the right to receive information) outweigh the costs to such an extent so as to find that the speech should, in principle, not be attributed to search engines—after all, autocomplete suggestions, for instance, are largely just holding a mirror up to society. Or are features such as these merely an unnecessary convenience?

Legislators and courts will increasingly have to deal with these issues, and regulation may be needed so as to provide guidance and more clearly identify to whom algorithmic speech should be attributed and how entities should be classified, as there will no doubt arise new forms of algorithmic speech that will further push these already ambiguous boundaries.⁹¹ Many questions remain, and, as will be seen

decide whether the text string that was searched for is true. However, having it discern between truth and falsity online would be extremely difficult. Even if it were to give greater weight to trusted sources or news organizations, the algorithm would likely have trouble if those sources published a mostly true story or a story about rumors, even if they were disproving them. Perhaps with time such algorithms may be developed so as to make this possible. *See also* Sean MacAvaney, Hao-Ren Yao, Eugene Yang, Katina Russell, Nazil Goharian & Ophir Frieder, *Hate Speech Detection: Challenges and Solutions*, 14 PLOS ONE, Aug. 20, 2019, at 1, 2 n.8 (2019) (noting challenges in detecting hate speech using machine learning techniques).

89. Whether or not they will actually be held liable requires further analysis. *See infra* Part II.B.2.

90. *Compare* Wannes Vandenbussche, Rethinking Non-Pecuniary Remedies for Defamation: The Case for Court-Ordered Apologies 22–23 (Aug. 22, 2018) (unpublished manuscript), <https://poseidon01.ssrn.com/delivery.php?ID=756021072024122096123084095004016069015032009051054004022004126025031121094099069078007052003023030014055091113106097093065126056022088032093121090124093078081001025053006012093083102000074085093105021074080087105112121089066027073016086064106124094064&EXT=pdf> [https://perma.cc/9458-SUSJ] (archived Aug. 20, 2020) (stating that in many countries across Europe, courts order retractions or apologies “in addition to or in lieu of monetary damages” (emphasis omitted)) *with* *Rogers v. Florence Printing Co.*, 106 S.E.2d 258, 263 (S.C. 1958) (“Retraction of a libel is matter to be considered in mitigation, but does not bar punitive damages in the absence of a statute so providing.”). The “innocent dissemination” defense in England would also be relevant here; for further information and how it fits with limited liability provisions internationally, see *infra* Part II.B.2.a.

91. One might look to copyright law’s standards on originality in relation to computer-generated works for guidance on how to attribute algorithmic speech. However, as these standards struggle to address newer forms of computer-generated works in the first place, it may prove difficult to extrapolate and apply to the situation at hand. Authorship of a work is largely dependent on the element of originality. In a case before the Court of Justice of the European Union (CJEU) involving a Danish computerized service that scanned various newspapers to produce 11-word extracts, the Court found that these snippets could satisfy the originality requirement of copyright so

in the following Parts, courts across Europe have struggled with how to examine search engine autocomplete suggestions, and it is far from settled.⁹²

long as they express the “author’s own intellectual creation.” Case C-5/08, *Infopaq International A/S v. Danske Dagblades Forening*, 2009 E.C.R. I-06569, ¶ 37. For most forms of algorithmic speech, it is difficult to imagine how the output could be the interacting user’s “own intellectual creation”, much less for responses from an autocomplete suggestion or from Microsoft Tay, given the programming behind the algorithm and the influence of the inputs of other users. *See id.*

The U.K. is one of a few jurisdictions that have a law that speaks more directly to computer-generated works. Section 9(3) of the Copyright, Designs and Patents Act (CDPA) 1988, c. 48 (U.K.), states:

“In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.”

This provision could provide guidance for one-to-one chat bots such as Microsoft’s Zo (the successor to Tay). Assuming the chat bot does not incorporate the inputs of other users—and it is unclear to what extent it does—it could be argued that the responses of Zo could be attributed to the user as they made the “arrangements necessary for the creation of the work.” However, the provision would still have problems with one-to-many chat bots, such as Tay, that incorporate the input of other users: how could the outputs of Tay be attributed to potentially thousands of (many times anonymous) users? While the spirit of the law appears to favor the programmer in cases of ambiguity in cases involving copyright, the Whitford Committee has stated that “the author of the output can be none other than the person, or persons, who devised the instructions and originated the data used to control and condition a computer to produce a particular result.” *See* WHITFORD COMMITTEE, COPYRIGHT AND DESIGNS LAW: REPORT OF THE COMMITTEE TO CONSIDER THE LAW ON COPYRIGHT AND DESIGNS, 1977, Cmnd. 6732, at ¶ 513 (UK).

Applying this to algorithmic speech, the determining factor in such cases may be the “particular result”, given that the responses of algorithms that produce speech are usually unpredictable. There is no doubt a need for further clarification on this issue, in both the context of copyright as well as the attribution of algorithmic speech. For a more in-depth analysis and comparison of the originality element in copyright law of computer-generated works in the U.K., Europe, and the U.S., see Andrés Guadamuz, *Do Androids Dream of Electric Copyright? Comparative Analysis of Originality in Artificial Intelligence Generated Works*, 2 INTELL. PROP. Q. 169 (2017).

92. Karapapa & Borghi, *supra* note 74, at 275–78. In one particularly interesting instance, an Italian court found that the autocomplete function makes the search engine neither an intermediary nor a content provider; it performs ‘active hosting’ and lies somewhere between the two. *See* Trib. di Milano, *Ordinanza*, 23 maggio 2013 (It.); *see also infra* Part II.B.2.a.

2. Liability for Algorithmic Speech

As alluded to above, the determination of who is a content provider informs the extent of liability. Content providers—those who provide their own content, adopt third-party content, exercise editorial control, or initiate the dissemination of third-party content—are typically held fully liable for the content that they publish.⁹³ Conversely, mere intermediaries enjoy limited liability, or in some instances, immunity, as they may not be aware of the exact content that they are transmitting.⁹⁴ As such, and given the foregoing discussion, this Part will focus on search engines and their algorithms that produce search results, and the extent to which they may limit their liability.

Limited liability frameworks may vary substantially by country, although there is some level of harmonization within the European Union.⁹⁵ A good, albeit broad, international definition of the principle of limited liability for intermediaries may be found in paragraph 2(a) of the *Joint Declaration on Freedom of Expression and the Internet*, which states:

No one who simply provides technical Internet services such as providing access, or searching for, or transmission or caching of information, should be liable for content generated by others, which is disseminated using those services, as long as they do not specifically intervene in that content or refuse to obey a court order to remove that content, where they have the capacity to do so ('mere conduit principle').⁹⁶

93. OSTER, MEDIA FREEDOM, *supra* note 36, at 58; OSTER, MEDIA LAW, *supra* note 36, at 14; Oster, *Liability of Intermediaries*, *supra* note 62, at 358. In some instances, content providers may also avoid liability as well through the use of defenses of honest opinion, publication on a matter of public interest, or parliamentary privilege. *See id.* at 351.

94. OSTER, MEDIA FREEDOM, *supra* note 36, at 58–59.

95.. This is largely due to the E-Commerce Directive, *infra* note 98. *See generally* IGNACIO GARROTE FERNÁNDEZ-DÍEZ, COMPARATIVE ANALYSIS ON NATIONAL APPROACHES TO THE LIABILITY OF INTERNET INTERMEDIARIES FOR INFRINGEMENT OF COPYRIGHT AND RELATED RIGHTS 67, 71–72 (World Intellectual Property Organization 2010); *see also World Intermediary Liability Map (WILMap)*, STAN. L. SCH. CTR. FOR INTERNET AND SOC'Y, <https://wilmap.law.stanford.edu/map> [<https://perma.cc/2K45-QY2T>] (archived Aug. 20, 2020).

96. Frank LaRue (U.N. Special Rapporteur on Freedom of Opinion & Expression), Dunja Mijatović (Organization for Security and Cooperation in Europe Representative on Freedom of the Media), Catalina Botero Marino (OAS Special Rapporteur on Freedom of Expression) & Faith Pansy Tlakula (ACHPR Special Rapporteur on Freedom of Expression & Access to Information), Joint declaration on freedom of expression and the Internet, OSCE (June 1, 2011), ¶ 2(a), <https://www.osce.org/files/f/documents/e/9/78309.pdf> [<https://perma.cc/7KC6-ASR8>] (archived Sept. 15, 2020) [hereinafter Joint Declaration].

Different forms of the “mere conduit” principle have become quite well established internationally, although there are some differences between the standards and their scope.⁹⁷

a. The European Framework

The European Union Directive 2000/31/EC (E-Commerce Directive) governs the liability of intermediaries within member states;⁹⁸ the results to be achieved are binding, but the forms and methods to reach those results are left to the national authorities.⁹⁹ Article 12 of the Directive protects intermediaries that are a “mere conduit,” and, although its usage in this provision is slightly different than that used above, the principle is largely the same.¹⁰⁰ The E-Commerce Directive further distinguishes intermediaries that perform “caching” and hosting functions in Articles 13 and 14, respectively, and the framework for liability differs between these three “classes” of intermediaries.¹⁰¹

Generally, these provisions provide protection from liability to intermediaries whose roles are “merely technical, automatic and passive,” and do not protect those that play “an active role of such a kind as to give [them] knowledge of, or control over, the data stored.”¹⁰² Within the confines of the E-Commerce Directive, this concept of data encompasses most illegal material, such as hate speech, child

97. See generally GARROTE FERNÁNDEZ-DÍEZ, *supra* note 95.

98. Council Directive 2000/31/EC, Directive on Electronic Commerce, 2000 O.J. (L 178) 1, 13 (EC) [hereinafter E-Commerce Directive].

99. Consolidated Version of the Treaty on the Functioning of the European Union art. 288, Jun. 7, 2016, 2016 O.J. (C 202) 47 [hereinafter TFEU].

100. E-Commerce Directive, *supra* note 98. Article 12 ‘Mere conduit’ states:

“1. Where an information society service is provided that consists of the transmission in a communication network of information provided by a recipient of the service, or the provision of access to a communication network, Member States shall ensure that the service provider is not liable for the information transmitted, on condition that the provider:

(a) does not initiate the transmission;

(b) does not select the receiver of the transmission; and

(c) does not select or modify the information contained in the transmission.

2. The acts of transmission and of provision of access referred to in paragraph 1 include the automatic, intermediate and transient storage of the information transmitted in so far as this takes place for the sole purpose of carrying out the transmission in the communication network, and provided that the information is not stored for any period longer than is reasonably necessary for the transmission.”

101. *Id.* art. 13–14.

102. *Id.*; see also *Delfi AS v. Estonia*, Eur. Ct. H.R. (2015) App. No. 64569/09, ¶ 52; Cases C-236/08 to C-238/08, *Google France and Google, Inc. v. Louis Vuitton Malletier SA*, *Google France v. Viaticum SA*, and *Luteciel S.A.R.L. and Google France v. Centre national de recherche en relations humaines (CNRRH) S.A.R.L.*, 2010 E.C.R. I-02417, ¶¶ 113–14, 121.

pornography, the infringement of intellectual property rights, and defamatory content.¹⁰³

The Court of Justice of the European Union (CJEU) is the principal judicial authority in the EU, which is tasked with ensuring the uniform application and interpretation of EU law,¹⁰⁴ and it has examined the intermediary liability provisions of the E-Commerce Directive on a number of occasions. A notable case in this regard involved the selling of goods on an online platform that infringed trademark rights. In *L'Oréal v. eBay*, the court found that the limitation of liability for hosting providers in Article 14(1) applies to an “operator of an online marketplace” so long as it “has not played an active role allowing it to have knowledge or control of the data stored.”¹⁰⁵ However, even if the platform has not played an “active role,” the limitation will not apply if the platform was “aware of facts or circumstances from which the illegal activity or information is apparent” and with that awareness it failed to act expeditiously to remove or disable access to the illegal content.¹⁰⁶ This reiterated the rule in *Google France v. Louis Vuitton*,¹⁰⁷ a case involving ads for counterfeit goods being shown when a trademarked term was entered in the search engine. There, the court found that “concordance between the keyword selected and the search term entered by an internet user is not sufficient of itself to justify the view that Google has knowledge of, or control over, the data entered into its system by advertisers and stored in memory on its server,” but also that “the role played by Google

103. See *First Report on the Application of Directive 2000/31/EC*, at 12, COM (2003) 702 final (Nov. 21, 2003) (“The limitations on liability provided for by the Directive are established in a horizontal manner, meaning that they cover liability, both civil and criminal, for all types of illegal activities initiated by third parties.”).

104. Consolidated Version of the Treaty on European Union art. 19, Oct. 12, 2012, 2012 O.J. (C 326) 13. The CJEU also considers the human rights enshrined in the Charter of Fundamental Rights of the European Union; however, this consideration is superficial in comparison with the analysis undertaken by the European Court of Human Rights when examining alleged violations of the European Convention on Human Rights brought before it. See *id.*

105. Case C-324/09, *L'Oréal SA v. eBay International AG*, 2011 E.C.R. I-6011, ¶ 123. An active role would be found where “it provides assistance which entails, in particular, optimising the presentation of the offers for sale in question or promoting them.” See *id.*

106. *Id.* ¶¶ 119–20, 124; E-Commerce Directive, *supra* note 98, art. 14(1)(a)–(b).

107. See Cases C-236/08 to C-238/08, *Google France and Google Inc. v. Louis Vuitton Malletier SA, Google France v. Viaticum SA, and Luteciel S.A.R.L. and Google France v. Centre national de recherche en relations humaines (CNRRH) S.A.R.L.*, 2010 E.C.R. I-02417, ¶ 120 (“Article 14 of Directive 2000/31 must be interpreted as meaning that the rule laid down therein applies to an internet referencing service provider in the case where that service provider has not played an active role of such a kind as to give it knowledge of, or control over, the data stored. If it has not played such a role, that service provider cannot be held liable for the data which it has stored at the request of an advertiser, unless, having obtained knowledge of the unlawful nature of those data or of that advertiser’s activities, it failed to act expeditiously to remove or to disable access to the data concerned.”).

in the drafting of the commercial message which accompanies the advertising link or in the establishment or selection of keywords is relevant.”¹⁰⁸

European Union member states must implement the E-Commerce Directive into their domestic legislation, which varies to some extent between the states. Even where the specific laws may appear different, the framework is still evident. For example, in the United Kingdom, the E-Commerce Directive was implemented through the Electronic Commerce (EC Directive) Regulations of 2002.¹⁰⁹ In cases that concern online defamation in England, this must be interpreted in coordination with other laws concerning defamation and the common law, which leads to some differences in terminology, yet the underlying principles function the same and in accordance with the E-Commerce Directive. Both content providers and intermediaries—if they knew or should have known of the defamatory content—can be considered publishers, who are then liable unless they are able to make a defense.¹¹⁰ Only intermediaries may avail themselves of the “innocent publication” defense,¹¹¹ which may be defeated if the intermediary fails to respond to a notice of complaint within an adequate time frame.¹¹² There are

108. *Id.* ¶¶ 117–18. The Court did not ultimately rule on this and left it for the national court. *Id.* ¶ 119.

109. Electronic Commerce (EC Directive) Regulations 2002, SI 2002/2013, arts. 17–19 (U.K.). Articles 17, 18, and 19 apply to mere conduits, caching providers, and hosting providers, respectively, and they are almost verbatim replicas of the provisions in the E-Commerce Directive. *See id.*

110. Whether intermediaries may be considered publishers, without knowledge of the defamatory content, is a point of contention. *Compare* Godfrey v. Demon Internet, Ltd. [1999] EWHC (QB) 244, [2001] [QB] 201 (Eng.) (Morland J) (ruling that under English common law that the ISP would clearly be the publisher of the content) *with* Bunt v. Tilley [2006] EWHC (QB) 407, [2007] 1 W.L.R. 1243 (Eng.) (Eady J) (finding that “an ISP which performs no more than a passive role in facilitating postings on the internet cannot be deemed to be a publisher at common law” unless he knew, or should have known when exercising reasonable care, that the publication was likely to be defamatory). Academics have also reached different conclusions on this matter. *Compare* Oster, *Liability of Intermediaries*, *supra* note 62, at 356–57 (arguing that it is illogical for an intermediary to become a publisher only after it becomes aware of the defamatory content—this is actually part of the innocent publication defense—and that intermediaries should be considered publishers if “they actively participate in making content known to another person, disregarding their knowledge of its defamatory character”; as such, it is a factual determination) *with* Karapapa & Borghi, *supra* note 74, at 272–73 (following a line of precedent including *Bunt v. Tilly* [confirmed in *Tamiz v. Google* [2013] EWCA (Civ) 68, [2012] EWHC (QB) 449 (Eng.)] to conclude that the considerations on whether an intermediary is a publisher change “after notification they knew or had reason to believe that their continued hosting of the materials caused, or contributed to, the publication of a defamatory statement”).

111. This is also known as the ‘innocent dissemination’ defense, and it corresponds to Article 14 of the E-Commerce Directive; in essence, this means that the intermediary took reasonable care in its publication and did not know nor have reason to believe that the publication was defamatory. *See* Defamation Act 1996, c. 31, § 1 (U.K.). It should be noted that some defenses are available to content providers as well. *See supra* note 93 and accompanying text.

112. *See* Defamation Act 2013, c. 26, § 5(3) (U.K.).

systems outside of the EU that utilize a similar framework to that found in the United Kingdom.¹¹³ Canada is one such country,¹¹⁴ as is Australia, where, in a case concerning defamatory connections in image and web results, Google was held liable for not removing the offending material within a reasonable time after it knew of the defamation complaint.¹¹⁵

While the CJEU has given guidance on the liability of certain online intermediaries, it has yet to provide more specific direction on how to analyze claims involving algorithmic speech. However, some national courts in Europe have had to interpret and apply the limited liability provisions of the E-Commerce Directive in cases involving a search engine's autocomplete function on a number of occasions, leading to different results. Most of these cases have concerned defamation, when a person or company's name was paired with unbecoming autocomplete suggestions. In a number of cases in Italy,¹¹⁶ France,¹¹⁷ and Germany,¹¹⁸ the provisions of the E-Commerce Directive (or rather their equivalent provisions in domestic law) were found not to apply to the autocomplete function, and thus the search engine (primarily Google) was held liable for the infringing content. However, in another instance, an Italian court found that Google was

113. This is likely due to the shared common law histories of the countries.

114. See Corey Omer, *Intermediary Liability for Harmful Speech: Lessons from Abroad*, 28 HARV. J.L. TECH. 289, 305–08 (2014).

115. See *Milorad Trkulja v Google Inc LLC (No 5)* [2012] VSC 533, No. 10096/2009 (Supreme Court of Victoria) (Austl.). Justice Beach of the Supreme Court of Victoria awarded the plaintiff \$200,000 in this case. See *id.*

116. See *AB v. Google*, Trib. Ordinario di Milano, 24 marzo 2011 (It.) (deciding that Google went beyond the mere hosting of user's searches when it automatically presented suggestions to other users based on conscious and commercially-based decision to do so); see also Trib. di Milano, *Ordinanza* 23 maggio 2013 (It.) (citing *C-324/09, L'Oréal SA v. eBay International AG*, 2011 E.C.R. I-6011, the court found that limitations of liability must be "limited to situations where the service provider is a mere intermediary of the information, totally unrelated to the content and thereby completely passive with respect to the content transmitted by third parties on the internet"; in doing so, the court coined the term 'active hosting', in other words, a service that lies between a content provider and mere intermediary, which may not utilize the E-Commerce hosting defense).

117. See *Mme C/Google France*, Tribunal de grande instance [TGI] [ordinary court of original jurisdiction] Montpellier, Oct. 28, 2010, (Fr.), *aff'd*, Cour d'appel [CA] [regional court of appeal] Montpellier, civ., Sept. 29, 2011 (Fr.). In this case, the plaintiff's name was associated with the title of an adult movie, and a similar conclusion was found to that in *AB v. Google*, Trib. Ordinario di Milano, 24 marzo 2011 (It.).

118. See BGH, May 14, 2013, VI ZR 269/12 (Ger.), <http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=64163&pos=0&anz=1> [<https://perma.cc/KQW5-2SPP>] (archived July 11, 2020) (finding that word associations, such as those resulting from an autocomplete suggestion, impart meaning and thus have the capability to be defamatory, and holding Google liable for not removing the false associations after receiving notification).

acting as a “caching” service within the meaning of the Directive,¹¹⁹ and still other courts in Italy,¹²⁰ Switzerland,¹²¹ and France¹²² held that the search engine is not liable based upon other reasons. As is apparent, and in order to achieve more uniformity, clear direction is needed in this area—whether through updated legislation or decisions by the CJEU.

The European Court of Human Rights (ECtHR) has also examined cases that concern the liability of intermediaries. In contrast to the CJEU, which interprets EU law, the ECtHR hears cases that involve allegations of violations of human rights perpetrated by a member state of the Council of Europe.¹²³ As such, it does not have the authority to rule on whether EU law was properly applied, as this lies within the competence of the CJEU. However, in two cases, the ECtHR has stated that additional factors should be taken into consideration regarding the liability of intermediaries in certain scenarios.¹²⁴

In one such case, *Delfi v. Estonia*, a news website was found to be liable for user comments posted in response to a news article, even after it took down the offending comments upon notification.¹²⁵ As this case was heard by the ECtHR as opposed to the CJEU, the court did

119. See *X v. Google*, Trib. Ordinario di Milano, N RG 2012/68306, 25 maggio 2013 (It.) (holding that Google is acting as a ‘caching’ service and thus may utilize the relevant defense in the E-Commerce Directive; as the autocomplete suggestions are based upon an algorithm, they are not under Google’s control within the meaning of being structured, organized or influenced).

120. See *X v. Google*, Trib. Ordinario di Pinerolo, 23 maggio 2012 (It.) (finding that autocomplete word associations may embody opinions and thus a defamatory meaning only where the statement is not true and the question was asked in a rhetorical or malicious manner).

121. See *Albert Tanneur Institut & Co v. Google Inc*, Tribunal Cantonal du Jura, 12 Feb. 2011 (Switz.) (imposing an obligation on search engines to remove autocomplete suggestions that may infringe one’s personality right would not be proportionate as it “would restrict inadmissibly the right to information” and would make continuing operation impossible in the face of potential liability).

122. See *Cour de Cassation [Cass.] [supreme court for judicial matters]*, 1e civ., Feb. 19, 2013, Bull. civ. I, No. 19 (Fr.) (*Pierre B. v. Google Inc.*); *Cour de Cassation [Cass.] [supreme court for judicial matters]* 1e civ., Jun. 19, 2013, Bull. civ. I, No. 625 (Fr.) (*Google v. Lyonnaise de garantie*). In both of these cases, the French Supreme Court found that an autocomplete function’s word associations are only a technical method to facilitate a search and are not expressions of opinion, which are necessary to base a claim of defamation. See *Cour de Cassation*, 1e civ., Feb. 19, 2013, Bull. civ. I, No. 19; *Cour de Cassation*, 1e civ., Jun. 19, 2013, Bull. civ. I, No. 625.

123. The Council of Europe is significantly larger than the European Union, and has 47 member states compared to the EU’s 28. For these 19 countries outside the scope of the EU, the ECtHR decisions on intermediary liability discussed below will have comparatively more importance on this issue than for their EU brethren. In addition, it should be reemphasized that the CJEU takes into account human rights issues in its rulings. See *supra* note 104.

124. See generally *Delfi AS v. Estonia*, Eur. Ct. H.R. App. No. 64569/09, (2015); *Magyar Tartalomszolgáltatók Egyesülete & Index.hu Zrt v. Hungary*, Eur. Ct. H.R. App. No. 22947/13, (2016) [hereinafter *MTE v. Hungary*].

125. See *Delfi AS v. Estonia* ¶ 162.

not rule on whether the limited liability provisions of the E-Commerce Directive were properly applied.¹²⁶ However, the court was satisfied that the news portal's exercise of control over the user comment section was such that it "went beyond that of a passive, purely technical service provider" and thus it played an "active role" within the meanings given to those concepts by *Google France v. Louis Vuitton* and *L'Oréal v. eBay*.¹²⁷ Further, the decision may also have implications for the future liability of intermediaries.¹²⁸ The court identified a number of aspects to be analyzed in order to determine whether holding the company liable for the comments posted by third-party users violates its freedom of expression, which include: "the context of the comments, the measures applied by the applicant company in order to prevent or remove defamatory comments, the liability of the actual authors of the comments as an alternative to the applicant company's liability, and the consequences of the domestic proceedings for the applicant company."¹²⁹ Ultimately, the court held

126. *Id.* ¶ 81. The proper forum for this case should have been the CJEU. See Oster, *Liability of Intermediaries*, *supra* note 62, at 360; OSTER, MEDIA LAW, *supra* note 36, at 238–39.

127. *Delfi AS v. Estonia* ¶¶ 52–53, 146; see also Cases C-236/08 to C-238/08, *Google France and Google Inc. v. Louis Vuitton Malletier SA*, *Google France v. Viaticum SA*, and *Luteciel S.A.R.L. and Google France v. Centre national de recherche en relations humaines (CNRRH) S.A.R.L.*, 2010 E.C.R. I-02417, ¶¶ 113–14, 120–21; Case C-324/09, *L'Oréal SA v. eBay International AG*, 2011 E.C.R. I-6011, ¶¶ 113, 116, 123.

128. While the court limited the scope of the decision so as to not include Internet discussion forums like bulletin boards, social media platforms that do not offer their own content, or content providers that are persons running a blog as a hobby, the line drawn is tenuous at best. There is little reason for Internet news portals that "provide for economic purposes a platform for user-generated comments on previously published content" to be distinguished from social media platforms. This implies that social media platforms are apparently able to escape liability solely because they did not publish content and invite users to comment—they merely let users publish content and invite other users to comment for economic purposes. If anything, social media networks such as Facebook would have more capacity (both economically and technologically) to be able to actively monitor illegal content through the use of algorithms and/or employees than a small local newspaper. See *Delfi v. Estonia* ¶¶ 115–16; see also *infra* note 129 and accompanying text; Ingrid Lunden, *Facebook to Add 3,000 to Team Reviewing Posts with Hate Speech, Crimes, and Other Harming Posts*, TECHCRUNCH (May 3, 2017), <https://techcrunch.com/2017/05/03/facebook-to-hire-3000-to-review-posts-with-hate-speech-crimes-and-other-harming-posts/> [<https://perma.cc/3KWX-6DRU>] (archived Aug. 21, 2020). Cf. *MTE v. Hungary*, *supra* note 124, ¶¶ 64, 91, where, in a situation very similar to *Delfi v. Estonia*, it was found that a news portal should not have been held liable for the comments of its users; the distinguishing difference was that the comments concerned did not constitute unlawful speech, such as hate speech or incitement to violence. The Court also reiterated the importance of inviting users to comment for "economic benefit" and found MTE's status as a self-regulatory non-profit body to be a contributing factor in its analysis.

129. *MTE v. Hungary*, *supra* note 124, ¶ 142. Some commentators have found that the implications of the Court's decision troubling from a freedom of expression standpoint, in that "(1) takedown upon notice is insufficient to avoid liability, (2) there may be some affirmative duty to monitor user-generated content, and (3) permitting

that there had been no violation of Defli's freedom of expression through the imposition of liability.

However, in *MTE v. Hungary*, which contained a similar factual scenario to that in *Delfi*,¹³⁰ the ECtHR appeared to take a step back from the earlier judgment and found that there was a violation of freedom of expression in holding the intermediaries liable.¹³¹ In making its decision, the court took into account the context and content of the impugned comments, the liability of the actual authors of the comments, the measures taken by the applicants and the conduct of the injured party, the consequences of the comments for the injured party, and the consequences for the applicants.¹³² The determining factor—and the primary difference with *Delfi*—was that the comments did not rise to the level of hate speech or incitement to violence.¹³³

As one can see, the analysis for intermediary liability may differ substantially depending on the court that is hearing the case due to the scope of judicial review. While the proper forum for cases involving intermediary liability under the E-Commerce Directive is the CJEU, it remains to be seen if, or to what extent, the CJEU will adopt some of the factors proposed by the ECtHR.

b. The United States and Elsewhere

The European framework can be contrasted with that found in the United States. The United States extends the “mere conduit” principle beyond the access provider level to both the hosting and content

anonymous posting should count against an intermediary's immunity.” See Omer, *supra* note 114, at 313–14 (citing Gabrielle Guillemin, *Case Law, Strasbourg: Delfi AS v Estonia: Court Strikes Serious Blow to Free Speech Online*, INFORM'S BLOG (Oct. 15, 2013), <https://inform.wordpress.com/2013/10/15/case-law-strasbourg-delfi-as-v-estonia-court-strikes-serious-blow-to-free-speech-online-gabrielle-guillemin/> [https://perma.cc/ZJ7A-5MEZ] (archived Aug. 21, 2020)). Two of the presiding judges also took issue with the judgment of the Court, saying that “all comments will have to be monitored from the moment they are posted. As a consequence, active intermediaries and blog operators will have considerable incentives to discontinue offering a comments feature, and the fear of liability may lead to additional self-censorship by operators.” See *Delfi v. Estonia* ¶ 1 (Sajó, J. and Tsotsoria, J., dissenting). However, not all commentators have found the decision so disturbing. See OSTER, MEDIA LAW, *supra* note 36, at 240–41 (arguing that the “collateral censorship” warned of in the dissenting opinion of Judges Sajó and Tsotsoria may be overblown because the Court limited the scope of the judgment, it did not impose an obligation on news portals to censor content in their comment sections, and the fact that other interests and rights should be accounted for, such as the right to reputation, which in the case of anonymous comments can only be corralled by regulating “those who control the code.”).

130. In this case, a nonprofit self-regulatory organization of Internet content providers (MTE) and an Internet news portal (Index.hu) were held liable by Hungarian courts for offensive comments posted on their websites by users. See generally *MTE v. Hungary*, *supra* note 124.

131. See generally *id.*

132. See *id.* ¶¶ 71–88.

133. See *id.* ¶¶ 64, 70, 91.

provider levels (which are many times one and the same).¹³⁴ Section 230 of the Communications Decency Act details that “[n]o provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider,” and as such, they are protected from the possibility of liability for this content.¹³⁵ Furthermore, intermediaries will not be liable for “any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected.”¹³⁶ This stands in stark contrast to the European framework discussed in the previous section, where an intermediary may become a content provider through the exercise of editorial control, thus exposing themselves to more liability.¹³⁷ Matters concerning intellectual property in the United States are governed by the Digital Millennium Copyright Act, the provisions of which are more similar to the limited liability provisions for host providers in the E-Commerce Directive.¹³⁸ This “safe harbor” provision protects hosting intermediaries from liability, so long as they were not aware of an infringement, or once they became aware or were notified by a third party, they “expeditiously” took down or disabled access to the infringing content; further, the intermediary may not financially benefit from the infringing activity directly.¹³⁹ A number of countries

134. Cf. E-Commerce Directive, *supra* note 98, art. 14 (stating that the hosting provider will not be liable if they do not have actual knowledge of the illegal content, or once they do have actual knowledge, they act quickly to remove access to the content).

135. Communications Decency Act of 1996, Pub. L. No. 104-104, sec. 509, § 230(c)(1), 110 Stat. 56, 138 (1996) (amending 47 U.S.C. § 230(c)(1) (2018)). Notably, this limited liability provision does not apply to matters that concern intellectual property law. Interestingly, the U.S. has been exporting this exception through free trade agreements with other countries including Chile, Colombia, and Morocco. See GARROTE FERNÁNDEZ-DÍEZ, *supra* note 95, §§ V, XI, XII.

136. Communications Decency Act sec. 509, § 230(c)(2)(a); see also *Langdon v. Google, Inc.*, 474 F.Supp.2d 622, 630 (D. Del. 2007). Google and Microsoft argued that their exercise of editorial discretion in filtering and deleting content from their services is protected by this provision. One might argue that this appears to leave the door wide open for private censorship with little to no transparency and often no appeal process; for instance, this enables Facebook to block constitutionally-protected speech such as nude pictures for merely violating Facebook’s terms of service. On the other hand, it encourages self-regulation. See *id.*

137. See Case C-291/13, *Sotiris Papasavvas v. O Fileleftheros Dimosia Etairia, Ltd.*, ECLI:EU:C:2014:2209, 11 Sept. 2014. Regarding a newspaper that operated a free online version of its articles, the CJEU found that because the newspaper had, in principle, knowledge about the information it posted and exercised control over that information, it could not be considered an intermediary within the understanding of Articles 12 to 14 of the E-Commerce Directive. See also E-Commerce Directive, *supra* note 98, and discussion on content providers versus intermediaries, Part II.B.1.

138. Digital Millennium Copyright Act, Pub. L. No. 105-304, sec. 202, § 512(c)(1), 112 Stat. 2860, 2879–80 (1998).

139. See *id.*

in Latin America have begun implementing similar frameworks, as the United States has been “exporting” it through free trade agreements.¹⁴⁰

Obviously, however, there are still other jurisdictions with different systems in place. Brazil has a system most similar to that posed in the Declaration quoted at the beginning of the section.¹⁴¹ Brazil’s Law no. 12.965/2014 (also known as the “Brazilian Internet Bill of Rights”) provides that Internet application providers should only be held liable if they fail to comply with a *judicial* order to disable access to the identified content.¹⁴² At the other end of the spectrum lies China, which has in the past used a strict liability regime for intermediaries, although it should be noted that its framework has been undergoing change so as to make it more similar to systems found abroad.¹⁴³

140. See J. Carlos Lara Gálvez & Alan M. Sears, *The Impact of Free Trade Agreements on Internet Intermediary Liability in Latin America*, in THE OXFORD HANDBOOK OF ONLINE INTERMEDIARY LIABILITY 172–73 (Oxford University Press, 2020).

141. See *supra* note 96 and accompanying text.

142. See Presidência da República, Lei No. 12.965, 23 de Abril de 2014 (Braz.). Article 19 states that “[i]n order to ensure freedom of expression and to prevent censorship, internet application providers may only be held civilly liable for damage resulting from content generated by third parties if after specific judicial order the provider fails to take action to make the content identified as offensive unavailable on its service by the stipulated deadline, subject to the technical limitations of its service and any legal provisions to the contrary.” It further states: “§ 1º The judicial order mentioned in the heading must contain, under the penalty of nullification, clear and specific identification of the content claimed to be a violation, which allows for the unequivocal identification of the content. § 2º In cases where there is an infringement on copyright laws and other related rights, this Article shall be applicable when specific legal precaution has been utilized, with full respect for freedom of expression and other guarantees provided for in Art. 5 of the Federal Constitution.” Requiring judicial orders can be viewed as more protective of freedom of expression as intermediaries will not err on the side of caution and take down content merely upon receiving notification of allegedly infringing content by a private party. On the other hand, the person or entity may suffer more damage by having the infringing content remain online until a court order is issued. This issue is addressed to an extent by Article 21, which establishes subsidiary liability for intermediaries that are disseminating private content (which may be images, videos or other material containing nudity or sexual acts) created by a third party; upon notification by the participant or legal representative, the intermediary must stop promotion, in a diligent manner, and within the scope and technical limits of the service, make said content unavailable. See *id.* arts. 19, 21.

143. See Qian Tao, *Legal framework of online intermediaries’ liability in China*, 14 DIGITAL POL’Y REG. GOVERNANCE 59, 61–62, 68 (2012); E-Commerce Law of the People’s Republic of China (adopted Aug. 31, 2018, effective Jan. 1, 2019), art. 38, <https://npcobserver.com/lawlist/e-commerce-law/> [https://perma.cc/5R53-NU3F] (archived Aug. 21, 2020).

c. Limited Liability Generally

At the risk of drawing too broad of generalizations from these disparate systems, generally, two conditions must be considered when an intermediary claims a limited liability defense: the intermediary's knowledge of the infringement, and the intermediary's response to a takedown notice of the claimed infringement.¹⁴⁴

Where then does this leave the creators or controllers of the algorithms used in search engines' autocomplete functions or in the search engine results themselves?

Regarding the autocomplete functions, the general trend is that search engines may not invoke the limited liability defense in the first place; granted, there is still no consensus on this point.¹⁴⁵ On its face, the inability to invoke the defense appears to be a reasonable approach. While this form of algorithm may not fall neatly into either content provider or intermediary categories, the autocomplete suggestions appear to be performing a function that is beyond that of a mere intermediary, and, in the end, the search engine has (editorial) control over the output as it has the capability to filter out certain results.

On the other hand, with search engine results, the proprietors are generally able to make use of limited liability defenses, assuming they respond to notifications of infringing content within a reasonable time.¹⁴⁶

There are also practical aspects that should be considered as well. Holding search engines liable for search results may lead to an unviable business model. Web search has become such an integral part of the Internet experience that it is difficult to imagine how one would find information online without the use of search engines, and the right to receive information would no doubt be implicated. Hence, the limitation of liability appears warranted.

However, autocomplete suggestions in their current state are not nearly as integral as search engine results are, and may be viewed as merely a convenience. Perhaps a taming of the feature is warranted, or maybe the benefits of even more easily accessible information outweigh the detriments so as to influence legislators to extend limited liability

144. There is arguably a third condition that must be considered: the intermediary's possible financial benefit from the relevant activity. However, there is no consensus in Europe on this point. See *Delfi AS v. Estonia*, Eur. Ct. H.R. App. No. 64569/09, ¶¶ 112–13 (2015).

145. See *supra* text accompanying notes 114–20.

146. What is a reasonable time will depend on the jurisdiction, and perhaps the nature of the infringing content, as seen above. See, e.g., Carolyn S. Toto and Kimberly Buffington, *The Complicated Relationship between DMCA Takedown Notices and the Word "Expeditious,"* INTERNET & SOC. MEDIA LAW BLOG (Jan. 19, 2016), <https://www.internetandtechnologylaw.com/the-complicated-relationship-between-dmca-takedown-notices-and-the-word-expeditious/> [https://perma.cc/Z9HP-2N44] (archived Sept. 21, 2020) ("For now, the relevant cases suggest that while responding within days or even weeks may be expeditious, taking months to respond is not.").

protections to the use of such algorithms, especially when the autocomplete function is essentially a reflection of others' searches.¹⁴⁷

Further regulation or guidance is clearly needed in this arena in order to clarify and harmonize the existing standards. The Internet, and the technology underpinning it, such as algorithms, has evolved in such a way so as to render many of the currently relevant provisions problematic to apply. Furthermore, algorithms and their speech outputs will only increase in complexity with time. The difficulty will no doubt lie in making the definitions and provisions technologically neutral so as to be able to adapt to changing technologies, while still being specific enough so as to be interpretable by courts when presented with unforeseen cases.

III. TO WHAT EXTENT IS ALGORITHMIC SPEECH WORTHY OF PROTECTION UNDER INTERNATIONAL STANDARDS OF FREEDOM OF EXPRESSION?

Given the foregoing discussion, the next question becomes: To what extent is algorithmic speech worthy of protection under international freedom of expression standards? Broadly, the following sections will examine to what extent interferences with algorithmic speech may be justified. First, possible interferences with algorithmic speech will be discussed.

A. *Interferences with Algorithmic Speech*

Interferences with or limitations on freedom of expression may be found, *inter alia*, through legislation, administrative regulations, or judicial decisions, which can prohibit or impose civil or criminal liability for a publication.¹⁴⁸

147. See discussion *supra* Part II.B.1 on attribution.

148. See Article 10(2) of the European Convention on Human Rights, which refers to different aspects that can constitute interference, such as "formalities, conditions, restrictions or penalties." ECHR, *supra* note 31, art. 10(2). It should also be noted that an interference with or limitation on freedom of expression does not mean that there has been a violation, and is often not a point of contention between the parties.

International courts have found interferences with freedom of expression in a wide variety of instances, such as through:

- an injunction preventing publication,¹⁴⁹
- a fine on a company for a user expressing hate speech or incitement to violence on a website,¹⁵⁰
- a criminal conviction and sentence for disseminating hate speech,¹⁵¹
- an order to disclose a source with a fine for refusing to do so,¹⁵²
- a criminal conviction and fine for defaming a religious group,¹⁵³
- a criminal penalty for genocide denial,¹⁵⁴
- a criminal conviction for publishing classified material,¹⁵⁵ the denial of a court to release a complaint pending before it,¹⁵⁶
- restricting the discussion of certain topics by the media,¹⁵⁷
- the overbroad blocking of websites,¹⁵⁸
- the prohibition on publishing taxation data by a data protection agency,¹⁵⁹ a statutory prohibition of paid political advertising on radio and television,¹⁶⁰
- denying a license to a television broadcasting company that is critical of the government,¹⁶¹ and additional liability for publishing something online in addition to print.¹⁶²

149. *Sunday Times v. United Kingdom* (No. 1), Eur. Ct. H.R. App. No. 6538/74, ¶ 49 (1979).

150. *See Delfi AS v. Estonia*, Eur. Ct. H.R. App. No. 64569/09, ¶ 118 (2015). In this case an interference was found with the applicant news portal's freedom of expression even though they were being fined for a comment posted by one of its readers. *See id.*

151. *Jersild v. Denmark*, Eur. Ct. H.R. App. No. 15890/89, ¶ 27 (1994).

152. *Goodwin v. the United Kingdom*, Eur. Ct. H.R. App. No. 17488/90, ¶ 28 (1996).

153. *Giniewski v. France*, Eur. Ct. H.R. App. No. 64015/00, ¶ 9 (2006).

154. *See Perinçek v. Switzerland*, Eur. Ct. H.R. App. No. 27510/08, ¶ 49 (2015).

155. *Stoll v. Switzerland*, Eur. Ct. H.R. App. No. 69698/01, ¶ 15 (2007).

156. *Társaság A Szabadságjogokért v. Hungary*, Eur. Ct. H.R. App. No. 37374/05 ¶ 15 (2009).

157. *Manole v. Moldova*, Eur. Ct. H.R. App. No. 13936/02, ¶ 33 (2009).

158. *Yıldırım v. Turkey*, Eur. Ct. H.R. App. No. 3111/10, ¶ 55 (2012). In this case, the applicant operated a blog on Google Sites, the entirety of which had been blocked due to an order of a court in a case that did not concern the applicant in any way. *See id.*

159. *See Satakunnan Markkinapörssi Oy v. Finland*, Eur. Ct. H.R. App. No. 931/13, ¶ 122 (2015).

160. *Animal Defenders International v. United Kingdom*, Eur. Ct. H.R. App. No. 48876/08, ¶ 78 (2013).

161. *See Granier (Radio Caracas Television) v. Venezuela*, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 293, ¶ 199 (June 22, 2015).

162. *Times Newspapers Ltd. v. United Kingdom* (Nos. 1 & 2), Eur. Ct. H.R. App. Nos. 3002/03 & 23676/03, ¶ 37 (2009). Having a new cause of action accrue every time defamatory material is accessed was known as the "Internet publication rule" in the U.K.; this can be contrasted with the "single publication rule" found in the U.S., which stipulates that the limitation period runs from when content is first published, whether in print or online. *Id.* ¶¶ 13, 24–25. It should be noted that the U.K. enacted a version of

Some of these examples would not currently apply to algorithmic speech, such as those pertaining to licensing or orders to disclose a source with a fine for refusing to do so. However, without legislation that specifically affects algorithms—whether directly or indirectly—interferences with freedom of expression that is exercised through algorithmic speech would largely operate no differently than they currently do when the means is more ordinary.¹⁶³ Thus, an interference with speech emanating from an algorithm could be found primarily through *ex post* measures such as when civil or criminal penalties for defamation, hate speech, or incitement to violence are imposed on the person or entity in control of the algorithm (i.e., the content provider or publisher).¹⁶⁴

Ex ante measures could also feasibly interfere with algorithmic speech, although due to their intrusiveness they are typically less preferable to *ex post* measures.¹⁶⁵ With the recent push for algorithmic accountability,¹⁶⁶ there is an increasing possibility that regulations will require algorithmic audits or transparency obligations,¹⁶⁷ and the

the single publication rule in section 8 of its Defamation Act. *See* Defamation Act 2013, c. 26, § 5(3) (U.K.).

163. By ordinary means, I mean to say through speech that is spoken or written by a person or entity, whether in print or online.

164. *See supra* Part II.B.2 for a more thorough discussion on liability.

165. *See* OSTER, MEDIA FREEDOM, *supra* note 36, at 144.

166. *See* Bryce W. Goodman, A Step Towards Accountable Algorithms?: Algorithmic Discrimination and the European Union General Data Protection 1–7 (2016) (unpublished manuscript) (on file with author) (discussing the extent to which the General Data Protection Regulation [GDPR] requires in regards to algorithmic audits, and what they might look like); Margot E. Kaminski & Gianclaudio Malgieri, *Multi-Layered Explanations from Algorithmic Impact Assessments in the GDPR*, in PROCEEDINGS OF ACM FAT* CONFERENCE (FAT* 2020) 68, 68–77 (2020) (examining how individual rights and systemic governance in the GDPR may allow for algorithms that are more accountable and explainable); *Algorithmic Accountability and Transparency in the Digital Single Market: Video of European Parliament Hearing*, MARIETJE SCHAAKE (Nov. 10, 2016), <https://marietjeschaake.eu/en/algorithmic-accountability-and-transparency-in-the-digital-single-market> [<https://perma.cc/8NXY-LMTX>] (archived July 11, 2020); Hemant Taneja, *The Need for Algorithmic Accountability*, TECHCRUNCH (Sept. 8, 2016), <https://techcrunch.com/2016/09/08/the-need-for-algorithmic-accountability/> [<https://perma.cc/QQ4B-4T4Q>] (archived July 11, 2020) (arguing for the self-regulation of algorithms through the use of open and transparent algorithmic watchdogs). The new General Data Protection Regulation has also given a nod to algorithmic discrimination in Recital 71, which states a requirement to “implement technical and organizational measures” that “prevent, inter alia, discriminatory effects on natural persons on the basis of racial or ethnic origin, political opinion, religion or beliefs, trade union membership, genetic or health status or sexual orientation, or that result in measures having such an effect.” Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 14 [hereinafter GDPR].

167. *See, e.g.*, EUR. PARLIAMENTARY RESEARCH SERV., A GOVERNANCE FRAMEWORK FOR ALGORITHMIC ACCOUNTABILITY AND TRANSPARENCY 3 (2019); Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. (2019).

required changes that may result from such audits could interfere with the algorithmic controller's freedom of expression. However, many questions remain as to how such a system would be implemented, and thus it is beyond the scope of this Article to explore in depth or hypothesize as to how algorithmic speech may be interfered with in the future through such regulations.¹⁶⁸

The following subparts will examine how an international court might analyze an interference with freedom of expression, where the speech involved is the output of algorithms, in order to determine whether there has been a violation.

B. Under What Circumstances Would Interferences with Algorithmic Speech be Justified?

For an interference with a fundamental right—such as freedom of expression—to be permissible, it must meet the requirements of the three-part test: the restriction must be prescribed by law, pursue a legitimate aim, and be necessary in pursuit of that aim.¹⁶⁹ The interfering state bears the burden to show that the limitation is justified.¹⁷⁰

This justification standard is well recognized around the world. Both Article 52 of the Charter of Fundamental Rights of the European Union and Article 10 § 2 of the European Convention on Human Rights include slightly different formulations of the three-part test,¹⁷¹ and the

168. See Goodman, *supra* note 166, at 6 (observing a number of outstanding questions such as who will be responsible for performing the audits, who will bear the cost of the audits, and how much will companies be expected to assist with the audits).

169. This test is the same in relation to freedom of expression on the Internet, where it has been stated that “any restriction that can affect this right must be provided for by law in the clearest and most precise terms possible, pursue a legitimate aim recognized by international law, and be necessary to accomplish that objective. . . .” Catalina Botero Marino (Special Rapporteur for Freedom of Expression), *Freedom of Expression and the Internet*, ¶ 122, O.A.S. Doc. OEA/Ser.L/V/II CIDH/RELE/INF.11/13 (Dec. 31, 2013) [hereinafter Botero Marino, *Freedom of Expression and the Internet*].

170. See, e.g., U.N., Econ. & Soc. Council, The Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights, UN Doc. E/CN.4/1985/4, annex, at 3 (1984) [hereinafter Siracusa Principles]; Human Rights Comm’n, *Coleman v. Australia*, Communication No. 1157/2003, U.N. Doc. CCPR/C/87/D/1157/2003, ¶ 7.3 (2005); Catalina Botero Marino (Special Rapporteur for Freedom of Expression), *Inter-American Legal Framework Regarding the Right to Freedom of Expression*, ¶ 83, O.A.S. Doc. CIDH/RELE/INF.2/09 (Dec. 30, 2009) [hereinafter Botero Marino, *Inter-American Legal Framework*].

171. Article 52(1) of the CFREU states:

Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others.

CFREU, *supra* note 33, art. 52.

European Court of Human Rights has repeatedly recognized this standard,¹⁷² as has the Inter-American Court on Human Rights.¹⁷³ Additionally, limitations on freedom of expression are to be interpreted strictly.¹⁷⁴

Despite the fact that these standards are well established, when examining algorithmic speech, we are in largely uncharted territory. The following subparts will therefore analyze interferences with various forms of algorithmic speech within the three-part test.

1. Prescribed by Law

Any restriction on freedom of expression, and by extension algorithmic speech, must be prescribed by law. First, this entails that the law that limits expression is accessible and foreseeable, the latter of which means that it must be sufficiently precise to allow persons to anticipate the consequences of their actions.¹⁷⁵ Even vaguely worded laws can pass this test,¹⁷⁶ although the degree of foreseeability will depend on a number of factors.¹⁷⁷

Article 10 § 2 of the ECHR states:

The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

ECHR, *supra* note 31, art. 10.

172. See *e.g.*, Müller v. Switzerland, Eur. Ct. H.R. App. No. 10737/84, ¶¶ 29–31 (1988); Aleksey Ovchinnikov v. Russia, Eur. Ct. H.R. App. No. 24061/04, ¶ 44 (2010).

173. See, *e.g.*, Usón Ramírez v. Venezuela, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 207, ¶ 88 (Nov. 20, 2009); Kimel v. Argentina, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 177, ¶ 58 (May 2, 2008); Canese v. Paraguay, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 111, ¶ 96 (Aug. 31, 2004); Herrera-Ulloa v. Costa Rica, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 107, ¶¶ 120–23 (July 2, 2004).

174. See EUR. CT. OF HUM. RIGHTS RESEARCH DIV., INTERNET: CASE-LAW OF THE EUROPEAN COURT OF HUMAN RIGHTS 17 (2015).

175. See Human Rights Comm'n, General Comment No. 34: Article 19: Freedoms of Opinion and Expression, ¶ 6, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011); Sanoma Uitgevers BV v. Netherlands, Eur. Ct. H.R. App. No. 38224/03, ¶ 81 (2010); Usón Ramírez v. Venezuela, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 207, ¶ 55 (Nov. 20, 2009); Cirio v. Uruguay, Case 11.500, Inter-Am. Comm'n H.R., Report No. 124/06, OEA/Ser.L/V/II.127, doc. 4 rev. 1 ¶ 64 (2006); Botero Marino, *Inter-American Legal Framework*, *supra* note 170, ¶ 69.

176. See Sunday Times v. United Kingdom (No. 1), Eur. Ct. H.R. App. No. 6538/74, ¶ 49 (1979).

177. See OSTER, MEDIA FREEDOM, *supra* note 36, at 113 (“The degree of foreseeability depends on the content of the law, its area of application, and the number and status of those to whom it is addressed.”); see also Cantoni v. France, Eur. Ct. H.R. App. No. 17862/91, ¶ 35 (1996).

Second, the law must legally protect persons against arbitrary interferences by public authorities. The delineation of the scope of those authorities' powers in applying the law must be sufficiently clear, and those authorities must not be given *carte blanche*.¹⁷⁸ Third, the law must meet certain formal requirements, such as being enshrined in a statute or a binding judicial decision.¹⁷⁹ This also entails that the law must be duly enacted in accordance with proper legislative procedure.¹⁸⁰ Finally, criminal sanctions must meet the principle of strict legality according to Article 15 of the ICCPR, Article 7 of the ECHR, and Article 9 of the ACHR.¹⁸¹

The currently existing laws that may affect algorithmic speech, such as those that pertain to defamation, hate speech, or copyright, etc., are unlikely to run afoul of this portion of the three-part test, as many of the more contentious ones have been tested through judicial scrutiny in contexts that do not involve the use of algorithms. However, with recent pushes for algorithmic accountability mentioned above, new laws may be enacted that specifically target algorithms.¹⁸²

2. Pursuit of a Legitimate Aim

Any interference with freedom of expression must pursue a legitimate aim. These aims, which include respecting the rights of others or protecting national security or public health, are stated in Article 19(3) of the ICCPR, Article 10(2) of the ECHR, Article 13(2) of the ACHR, and Article 52 of the CFREU.¹⁸³ While there is some

178. See Human Rights Comm., *Pinkney v. Canada*, Communication No. 27/1978, U.N. Doc. CCPR/C/OP/1, at 95, ¶ 34 (1980); Human Rights Comm'n, General Comment No. 27: Freedom of Movement (article 12), ¶ 13, U.N. Doc. CCPR/C/21/Rev.1/Add.9 (Nov. 2, 1999). See also *Siracusa Principles*, *supra* note 170, at 4; *Sanoma Uitgevers BV v. Netherlands*, Eur. Ct. H.R. App. No. 38224/03, ¶ 82 (2010); *Telegraaf Media Nederland Landelijke Media BV v. Netherlands*, Eur. Ct. H.R. App. No. 39315/06, ¶ 102 (2012); *Reyes v. Chile*, Inter-Am. Ct. H.R. (ser. C) No. 151, ¶ 89 (Sept. 19, 2006); *Botero Marino*, *Inter-American Legal Framework*, *supra* note 170, ¶ 70.

179. See *Sanoma Uitgevers BV v. Netherlands*, Eur. Ct. H.R. App. No. 38224/03, ¶ 83 (2010); *Miloslavsky v. United Kingdom*, Eur. Ct. H.R. App. No. 18139/91, ¶ 37 (1995); *X., Ltd. & Y. v. United Kingdom*, Eur. Comm'n H.R. App. No. 8710/79, ¶ 7 (1982); *Sunday Times v. United Kingdom* (No. 1), Eur. Ct. H.R. App. No. 6538/74, ¶ 49 (1979).

180. See *Barthold v. Germany*, Eur. Ct. H.R. App. No. 8734/79, ¶ 48 (1985); The Word "Laws" in Article 30 of the American Convention on Human Rights, Advisory Opinion OC-6/86, Inter-Am. Ct. H.R. (ser. A) No. 6, ¶¶ 22–24 (May 9, 1986).

181. See *Khodorkovskiy & Lebedev v. Russia*, Eur. Ct. H.R. App. Nos. 11082/06 & 13772/05, ¶ 778–85 (2013); *Usón Ramírez v. Venezuela*, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 207, ¶ 55 (Nov. 20, 2009); *Botero Marino*, *Inter-American Legal Framework*, *supra* note 170, ¶ 72.

182. See *supra* note 166.

183. For example, Article 13(2) of the ACHR states:

The exercise of the right provided for in the foregoing paragraph shall not be subject to prior censorship but shall be subject to subsequent imposition of

variation between the aims listed—in particular with the provision in the ECHR—they are functionally the same.¹⁸⁴

Generally, these listed aims are exclusionary;¹⁸⁵ however, certain ideas have crept in, such as media pluralism, and others have been expanded upon through courts' jurisprudence, such as the "rights of others."¹⁸⁶ Courts review the claimed legitimate aims with strict scrutiny, and thus there is no margin of appreciation given by courts to impugned states regarding this prong.

Similar to the previous section, unless there are unforeseen developments, it is unlikely for there to be any issues with interferences to algorithmic speech based upon the state not pursuing a legitimate aim in that interference. Restrictions on speech through laws on defamation, hate speech, or copyright are without doubt pursuing a legitimate aim as the framework currently stands.¹⁸⁷ As will be seen in the following Parts, the questions posed may not be so easily dismissed and will require more analysis than these two elements just discussed.

liability, which shall be expressly established by law to the extent necessary to ensure:

- a. respect for the rights or reputations of others; or
- b. the protection of national security, public order, or public health or morals.

ACHR, *supra* note 32, art. 13(2). The equivalent provisions in the other documents are very similar. A basic understanding of what these different terms mean may be found in the Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant and Political Rights. *See* Siracusa Principles, *supra* note 170, at 4–6.

184. *See* OSTER, MEDIA FREEDOM, *supra* note 36, at 115.

185. *See* Human Rights Comm'n, General Comment No. 34: Article 19: Freedoms of Opinion and Expression, ¶ 22, U.N. Doc. CCPR/C/GC/34 (Sept. 12, 2011).

186. *See* OSTER, MEDIA FREEDOM, *supra* note 36, at 115.

187. *See* *Jersild v. Denmark*, Eur. Ct. H.R. App. No. 15890/89, ¶ 27 (1994).

3. Necessary in Pursuit of that Aim

Finally, any interference with freedom of expression must be necessary in pursuit of the legitimate aim in order to be justified.¹⁸⁸ This prong of the three-part test, which is also known as the principle of proportionality, is comprised of three elements itself: suitability, necessity, and proportionality *sensu stricto* (proportionality in the narrow sense).¹⁸⁹

Furthermore, and in contrast to the “pursuit of a legitimate aim” prong, states are afforded a margin of appreciation by courts that are reviewing interferences with freedom of expression.¹⁹⁰ This is because international courts are not intended to take the place of their national counterparts, but to review the decisions under the relevant human rights instrument.¹⁹¹ A margin of appreciation is applicable to all of the three steps that comprise the proportionality prong, albeit in different respects.¹⁹² As the first two steps—suitability and necessity—refer to factual matters, the margin of appreciation is for the empirical uncertainties of the reviewing court, whereas proportionality *sensu stricto* concerns a normative analysis, and thus the margin of appreciation addresses the court’s normative uncertainties, such as those regarding the morals or religion of the local people.¹⁹³

Suitability refers to the examination of whether the interference was in fact appropriate to achieve the legitimate aim.¹⁹⁴ In other

188. See *Novaya Gazeta & Borodyanskiy v. Russia*, Eur. Ct. H.R. App. No. 14087/08, ¶ 32 (2013); *Canese v. Paraguay*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 111, ¶ 95 (Aug. 31, 2004); *Herrera-Ulloa v. Costa Rica*, Preliminary Objections, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 107, ¶ 120 (July 2, 2004); see generally U. N., Econ. & Soc. Council, The Siracusa Principles on the Limitation and Derogation Provisions in the International Covenant on Civil and Political Rights, UN Doc. E/CN.4/1985/4, annex (Sept. 28, 1984).

189. See Human Rights Comm., *Ballantyne v. Canada*, Communication No. 359, 385/89, U.N. Doc. CCPR/C/47/D/359/1989 and 385/1989/Rev.1, ¶ 11.4 (1993); *Cirio v. Uruguay*, Merits, Case 11.500, Inter-Am. Comm’n H.R., Report No. 124/06, OEA/Ser.L/V/II.127, doc. 4 rev. 1 ¶ 65 (2006); *Canese v. Paraguay*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 111, ¶ 96 (Aug. 31, 2004); *Sunday Times v. United Kingdom* (No. 1), Eur. Ct. H.R. App. No. 6538/74, ¶ 62 (1979); *Botero Marino*, *Inter-American Legal Framework*, *supra* note 170, ¶¶ 85–88.

190. See OSTER, *MEDIA FREEDOM*, *supra* note 36, at 118.

191. See *Handyside v. United Kingdom*, Eur. Ct. H.R. App. No. 5493/72, ¶ 50 (1976); *Sunday Times*, Eur. Ct. H.R. App. No. 6538/74, ¶ 59. See also Article 2 of the Optional First Protocol to the ICCPR, Article 46(1)(a) of the ACHR, and Article 35(1) of the ECHR, which stipulate that domestic remedies must be exhausted before taking the case to an international court. ICCPR, *supra* note 30, art. 2; ACHR, *supra* note 32, art. 46(1)(a); ECHR, *supra* note 31, art. 35(1).

192. See OSTER, *MEDIA FREEDOM*, *supra* note 36, at 118.

193. See *id.* at 119–20.

194. See *Botero Marino*, *Inter-American Legal Framework*, *supra* note 170, ¶¶ 85–88; *Kimel v. Argentina*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 177, ¶¶ 68–71 (May 2, 2008); OSTER, *MEDIA FREEDOM*, *supra* note 36, at 124.

words, the interference must be effectively conducive to attaining the objective in question.¹⁹⁵

The second step, necessity, involves the determination of whether the interference used was the least intrusive measure among those which could achieve the legitimate aim;¹⁹⁶ within the European and Inter-American human rights systems, this is also known as “pressing social need.”¹⁹⁷ Where the analysis of the pursuit of a legitimate aim looks towards the “end,” this test focuses on the “means.”¹⁹⁸ The margin of appreciation given by the court is particularly important at this stage as it affects the level of scrutiny applied when examining the measure that was the cause of the interference.¹⁹⁹ In regard to algorithmic speech, this necessity test could definitely become a point of contention, especially if an invasive form of algorithmic auditing was implemented through regulation, for example.

Finally, proportionality *sensu stricto* refers to the analysis of whether the interference was strictly proportionate in relation to the legitimate aim that is pursued.²⁰⁰ It is essentially a balancing exercise, where in the present instance, freedom of expression must be weighed against the advantages of the interference.²⁰¹ The nature and severity of the interference is examined, as well as the importance of the competing rights or interests and the extent to which they are affected.²⁰² In general, the various human rights are deserving of equal respect.²⁰³

Unlike the previous two prongs of the three-part test that were examined, the “necessary in pursuit of the legitimate aim” (or proportionality) prong is more likely to be the subject of contention. In the following Parts, other rights and interests that might be implicated with algorithmic speech and its interference will be examined, and problems that international adjudicators may have when reviewing cases that involve algorithmic speech will be anticipated and analyzed.

195. See Botero Marino, *Inter-American Legal Framework*, *supra* note 170, ¶¶ 85–88.

196. See OSTER, MEDIA FREEDOM, *supra* note 36, at 124–25.

197. See *Handyside v. United Kingdom*, Eur. Ct. H.R. App. No. 5493/72, ¶ 49 (1976); *Fontevicchia and D’Amico v. Argentina*, Merits, Reparations, and Costs, Judgment, Inter-Am. Ct. H.R. (ser. C) No. 238, ¶ 54 (Nov. 29, 2011).

198. See OSTER, MEDIA FREEDOM, *supra* note 36, at 124–25.

199. See *id.*

200. See *id.* at 125. See also Botero Marino, *Freedom of Expression and the Internet*, *supra* note 170, ¶ 124.

201. See OSTER, MEDIA FREEDOM, *supra* note 36, at 125.

202. See *id.*

203. See *Axel Springer AG v. Germany* (No. 1), Eur. Ct. H.R. App. No. 39954/08, ¶ 87 (2012); *Delfi AS v. Estonia*, Eur. Ct. H.R. App. No. 64569/09, ¶ 118 (2015). But see *Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos (AEPD)*, ECLI:EU:C:2014:317, ¶ 81 (May 13, 2014).

a. Other Rights and Interests at Play

As just described, the court must take into account a number of rights and interests when there has been an interference with the right to freedom of expression. Regarding algorithmic speech in particular, there are several interests that might be relevant when analyzing whether there has been a violation of the right. It must be noted that some rights and interests may be relevant to certain forms of algorithmic speech and not others. In addition to the nature of the algorithm, the factual scenario, including the nature of the interference, will also dictate which interests are relevant.

With freedom of expression, there is both the right to impart information, as well as the right to receive information.²⁰⁴ When there has been an interference with algorithmic speech, the right to receive information on the part of the user may also be implicated.²⁰⁵ While this may not apply to the likes of more purely “social” interactions with Microsoft Tay, it would likely apply to search engine results, the algorithms that power more informational chat bots like Amazon’s Alexa, Apple’s Siri, Google’s Assistant, or Microsoft’s Cortana, and to a lesser extent a search engine’s autocomplete suggestions.²⁰⁶

If the interference is such that it imposes onerous obligations on the controller of the algorithm, and thus makes the business model difficult to sustain, the freedom to conduct a business might be impeded.²⁰⁷ Such obligations could occur through requirements to filter or monitor the output of the algorithm, or through requirements for algorithmic transparency and audits. This could potentially apply to most of the forms of algorithmic speech discussed above, such as search engine results and their autocomplete function as well as more advanced chat bots, but would be less likely to apply to news-producing algorithms or simple chat bots with standardized scripts or outputs. However, if there were a system in place that required algorithmic

204. *Times Newspapers, Ltd. v. United Kingdom* (Nos. 1 & 2), Eur. Ct. H.R. App. Nos. 3002/03 & 23676/03, ¶ 27 (2009); *Guerra v. Italy*, Eur. Ct. H.R. App. No. 14967/89, ¶ 53 (1998); *Thorgeir Thorgeirson v. Iceland*, Eur. Ct. H.R. App. No. 13778/88, ¶ 63 (1992); *Observer & Guardian v. United Kingdom*, Eur. Ct. H.R. App. No. 13585/88, ¶ 59 (1991).

205. See *Társaság A Szabadságjogokért v. Hungary*, Eur. Ct. H.R. App. No. 37374/05, ¶¶ 35–36 (2009).

206. BGH, May 14, 2013, VI ZR 269/12, <http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/document.py?Gericht=bgh&Art=en&nr=64163&pos=0&anz=1> [<https://perma.cc/KQW5-2SPP>] (archived July 11, 2020) (finding that there is no general obligation for the search engine to seek out possible infringements of the autocomplete algorithm).

207. See *Case C-70/10, Scarlet Extended v. SABAM*, 2011 E.C.R. I-11959, ¶¶ 47–49; *Case C-360/10, SABAM v. Netlog NV*, ECLI:EU:C:2012:85, ¶¶ 44–47 (Feb. 16, 2012). Although the context of these two cases were slightly different, in both of them it was found that an obligation to install and maintain a filtering system did not strike a fair balance between the protection of the intellectual property right of copyright holders and the freedom to conduct a business.

transparency in a form that did not sufficiently occlude the source code or underlying algorithm, and it were to leak, the competitive advantage gained through the use of that algorithm or code may invoke the freedom to conduct a business for a broader array of algorithms that produce speech.

A person's right to privacy or right to data protection may be infringed if the algorithmic output contains the data sensitive to one's private life.²⁰⁸ This is particularly relevant for search engines,²⁰⁹ but it may also become relevant for a search engine's autocomplete function or for informational chat bots.

Intellectual property rights could be implicated on both sides of the balancing equation. In the case of search engines, informational chat bots, and perhaps even autocomplete functions, algorithmic speech may involve the intellectual property rights of copyright holders if there are links to infringing content.²¹⁰ On the other hand, and related to the freedom to conduct a business mentioned above, algorithmic transparency could be implemented in such a way so as to breach the intellectual property rights of the creator or controller of the algorithm in the algorithm itself.

The application of the above rights and interests will no doubt depend on the factual scenario of the case at hand. As algorithmic speech, interferences, and other facts of a case may all vary widely, it is a bit difficult at this point in time to hypothesize how algorithmic speech may be interfered with and impacted. However, given the fact that the balancing exercise may take into consideration multiple rights and interests on both sides of the scale, it is particularly flexible in handling a wide range of situations.

b. Issues Going Forward

There are numerous issues that may affect or influence the foregoing analysis in the future. Many of these involve questions of policy and societal values that accompany granting algorithmic speech protection.

208. See Case C-131/12, *Google Spain SL v. Agencia Española de Protección de Datos (AEPD)*, ECLI:EU:C:2014:317, ¶ 81 (May 13, 2014). While this case was decided within the context of the CFREU, Article 12 of the UDHR, Article 17 of the ICCPR, Article 8 of the ECHR, and Article 11 of the ACHR contain similar principles. Further, the principle underlying the case, that of the 'right to be forgotten' or 'right to de-list', is having a global effect with countries around the globe considering legislation that would enshrine the right.

209. See *id.*

210. See Case C-160/15, *GS Media BV v. Sanoma Media Netherlands BV*, ECLI:EU:C:2016:644, ¶ 55 (Sept. 8, 2016); Case C-360/10, *SABAM v. Netlog NV*, ECLI:EU:C:2012:85, ¶¶ 44–47 (Feb. 16, 2012); Case C-70/10, *Scarlet Extended v. SABAM*, 2011 E.C.R. I-11959, ¶¶ 47–49.

The aforementioned potential for more stringent algorithmic accountability mechanisms will likely be one such issue that will heavily affect how algorithmic speech is viewed by courts. If the form of algorithmic accountability is such that it heavily interferes with the operations of the creator or controller of the algorithm, such as through the use of periodic and invasive audits, it could be prohibitively expensive and impede innovation. Nonetheless, this may be an adequate trade off—from a societal standpoint—in order to protect users from hate speech and discrimination.²¹¹ However, such an intrusive framework for accountability is unlikely to happen in the near term, as the industry will likely push for self-regulation to the extent possible,²¹² and current measures—such as the GDPR—do not currently prescribe anything quite so extreme.²¹³

Courts and legislators may also have to take into account the nature and importance of the service that is making use of the algorithm. Certain services may be worthy of carving out protections to protect them from liability so that they remain viable. Search engines are extremely important to the modern usage of the web and will likely remain so for the foreseeable future.²¹⁴ While perhaps not rising to the level of search engines, other forms of algorithmic speech also may provide substantial value to society. For example, algorithmically generated news stories may free up resources for newspapers to focus on topics that require more investigation, thus enabling better information dispersion to the public, and some chat bots can potentially save businesses money by partially automating support services.

Further, decisions taken by legislators and courts can have large and wide-ranging effects given the fact that the law surrounding this area is currently poorly defined. Without care, a decision directed

211. Discrimination is outside the scope of this article, but it is particularly relevant in the realm of algorithmic biases, especially those concerning loans and credit scores.

212. Sandra Wachter has stated that “[t]he industry fear is that [companies] will have to disclose their code[.]” Louise Matsakis, *What Does a Fair Algorithm Actually Look Like?*, WIRED (Oct. 11, 2018), <https://www.wired.com/story/what-does-a-fair-algorithm-look-like/> [<https://perma.cc/2DQX-RWDU>] (archived July 11, 2020) (alteration in original).

213. To combat algorithmic discrimination, the GDPR encompasses two forms of algorithmic accountability: data sanitization (Articles 9 and 22(2)), or “the removal of special categories from datasets used in automated decision making”, and algorithmic transparency (Articles 12, 13(2)(f), and 14(2)(g)), or the “right to explanation” which includes “meaningful information about the logic involved, as well as the significance and the envisaged consequences” when automated decision making or profiling takes place.” Goodman, *supra* note 166, at 2. Impact assessments (Article 35) have also been discussed as a means of implementing algorithmic accountability. Kaminski & Malgieri, *supra* note 166, at 68.

214. If it were to be replaced, it would likely be with another form of algorithm; perhaps users will increasingly shift towards informational chat bots with voice support, for instance.

towards Microsoft Tay, which has comparatively little value to society, will also likely affect chat bots providing customer service or informational chat bots like Google's Assistant and Apple's Siri. This can in turn have an effect on these important informational services, as they may "censor" themselves—or rather reduce the quality or accuracy of the information provided—in order to escape potential future liability.

Additionally, certain forms of algorithms that lie on the fringe of speech may present difficult situations for courts. For example, Facebook uses algorithms to curate news stories (as well as cat videos) that a user's friends post, a number of which are then displayed in the user's "feed." Does this constitute a form of speech? At first glance, it may appear clear that it does not. But what if the curation of new stories displayed to users is such that it favors one side, so as to push a certain viewpoint or create a filter bubble—perhaps to keep users engaged—could it then be said to be Facebook's speech?²¹⁵ After all, this could be seen as a form of editorial control.²¹⁶ It is, however, a private company, which may have its own opinion, and the profit motive is a strong incentive to keep its algorithms content-neutral so as to not disenfranchise users who may not share the same opinion. On the other hand, with an ever-increasing number of users, perhaps one day there may be a need for a reimagining of what constitutes a public space for speech.²¹⁷

Finally, from a broad perspective, there are benefits to protecting algorithmic speech, in addition to the services provided by the algorithms themselves. Perhaps most importantly, it protects innovation. If the creators or controllers of algorithms were held strictly liable for every offense committed by those algorithms, it may become unfeasible to continue creating or operating new algorithms, and the public would no longer be able to benefit from these creations. Additionally, this would disproportionately affect small companies and startups, which may not have the resources to pay fines or fight battles in court. By contrast, the risks are relatively small. People might be defamed or subject to harmful speech by interacting with algorithms that have not undergone a sufficient vetting process.²¹⁸ However, the legal framework for dealing with such speech should be able to

215. See Aarti Shahani, *From Hate Speech to Fake News: The Content Crisis Facing Mark Zuckerberg*, NPR (Nov. 17, 2016), <http://www.npr.org/sections/alltechconsidered/2016/11/17/495827410/from-hate-speech-to-fake-news-the-content-crisis-facing-mark-zuckerberg> [https://perma.cc/HMB6-Y62U] (archived July 11, 2020).

216. See *supra* Part II.B.1.

217. See Josh Constine, *Facebook Now Has 2 Billion Monthly Users...and Responsibility*, TECHCRUNCH (June 27, 2017), <https://techcrunch.com/2017/06/27/facebook-2-billion-users/> [https://perma.cc/RUM5-4ZE4] (archived July 11, 2020).

218. See, e.g., *supra* Part II.B.1 (discussing Microsoft Tay and attribution).

adequately address harmful algorithmic outputs as well as it does for person-to-person interactions for the time being, especially if relatively minor adjustments are made.²¹⁹

As algorithms increase in complexity, so too will the issues they present. While the current framework for freedom of expression is largely sufficient to be able to handle the problems of today, only time will tell whether it will be able to adequately adapt to the intricacies of the future.²²⁰

IV. CONCLUSION

As algorithms have become increasingly common, so have algorithms that approximate speech. Given this development, this Article analyzed various forms of algorithmic speech within the international freedom of expression framework. This Article focused on the algorithms that would fall within three categories, defined as such: curated production (e.g., news stories, search engine results), interactive/responsive production (e.g., simple chat bots), and semi-autonomous production (e.g., Microsoft's Tay, search engine's autocomplete function).

One of the first issues examined was to whom algorithmic speech should be attributed and whether the creators or controllers should be considered content providers or mere intermediaries, which ultimately has implications for liability. This determination turns on a number of factors, such as whether the algorithm is providing its own content—or adopting it as its own—and editorial control. While the determination may seem relatively straightforward, and it is in some cases, certain forms of algorithmic speech present problems. For instance, with the curated production of news stories, editorial control is no doubt exercised, and thus the controller of the algorithm would be a publisher. On the other hand, adaptive chat bots like Microsoft Tay and a search engine's autocomplete function both use the input of other users when formulating responses and suggestions, respectively. As such, they could be considered the speech of other users, rather than that of the creator or controller of the algorithm. For

219. In certain instances, it may even be superior due to the 'data trail' left behind. Furthermore, people can oftentimes be erratic or unpredictable and not abide by requests to stop communication. At least with algorithms, if they start to malfunction, one can avoid interacting with it in most instances and it will likely be taken offline quickly in order to avoid potential liability—or loss of reputation—as in the case of Microsoft Tay.

220. Indeed, if fully autonomous artificial intelligence is realized, fairly significant changes may have to be made to the current framework. Some authors have already proposed frameworks and rough timelines for rolling out the attribution of personhood to nonhuman computer agents. See, e.g., Bert-Japp Koops, Mireille Hildebrandt & David-Oliver Jaquet-Chiffelle, *Bridging the Accountability Gap: Rights for New Entities in the Information Society?*, 11 MINN. J.L. SCI. & TECH. 497 (2010).

interactive/responsive production such as these, even though there is user input, the output could very well be viewed as adopting user input under the “ordinary reasonable person” standard; hence the controllers lean towards being considered content providers. In the end, the creator or controller of the algorithm has final control over the output in most cases, and could thus be said to have editorial control. On the other hand, it is likely that an ordinary reasonable person would consider that search engine results are lists of content that are neither provided nor adopted by the search engine itself, as it only presents excerpts and links to content provided elsewhere. Thus, the search engine in this context is the most likely among the algorithms discussed to be a mere intermediary and to avail itself of the associated limitations of liability. More guidance from legislators and courts would benefit this area of law, especially looking forward.

The liability of the creators or controllers of algorithmic speech was also analyzed. As frameworks for limited liability differ around the world, a number of them were discussed and compared, and general principles were drawn from them. While the analysis for intermediary liability may differ substantially depending on the court that is hearing the case due to the scope of judicial review, very generally, two conditions must be considered when an intermediary claims a limited liability defense. The first is the intermediary’s knowledge of the infringement, and the second is the intermediary’s response to a takedown notice of the claimed infringement. The general trend in Europe is for search engines to be liable for autocomplete suggestions, which appears to be a reasonable approach. As for search engine results, the proprietors are generally able to make use of limited liability defenses, assuming they respond to notifications of infringing content within a reasonable time. Further regulation or guidance was found to be clearly needed in this area, so as to clarify and harmonize the existing standards. The Internet, and the technology underpinning it, such as algorithms, has evolved in such a way so as to make many of the currently relevant provisions ambiguous or inapplicable. Furthermore, algorithms and their speech outputs will only increase in complexity with time.

The Article then examined the extent to which algorithmic speech is worthy of protection under international standards of freedom of expression. Much will depend upon forms of algorithmic accountability that may be required in the future, as some forms could be considered invasive and burdensome so as to interfere with freedom of expression. Until then, if there is an interference with algorithmic speech, potential interferences with algorithmic forms of expression will be largely the same as they currently are for human speech, such as fines or criminal penalties.

The next Parts go on to investigate interferences under the three-part test. Any restriction or interference with freedom of expression must be prescribed by law, pursue a legitimate aim, and be necessary

in pursuit of that aim. The interfering state bears the burden to show that the limitation is justified. Being prescribed by law entails that the law that limits expression is accessible and foreseeable, the law must legally protect persons against arbitrary interferences by public authorities, and the law must meet certain formal requirements. Any interference with freedom of expression must also pursue one of the legitimate aims delineated in human rights instruments. Without a form of algorithmic accountability or another restriction targeting algorithms in place, the first two prongs—prescribed by law and pursuit of a legitimate aim—will again operate as they do for more typical forms of speech. In other words, where existing laws on defamation, hate speech, or copyright are used to restrict algorithmic forms of expression, they are unlikely to be successfully challenged.

As for the necessary in pursuit of a legitimate aim prong, the focus is on the proportionality *sensu stricto* step, which refers to the analysis of whether the interference was strictly proportionate in relation to the legitimate aim that is pursued. This involves a balancing exercise, where in the present instance, freedom of expression must be weighed against the advantages of the interference. The nature and severity of the interference is examined, as well as the importance of the competing rights or interests and the extent to which they are affected. In addition to freedom of expression, there are a number of rights and interests to consider when a court examines an interference with algorithmic speech. Among these would be the right to receive information (as the other side of the right to impart information), the freedom to conduct a business, the right to privacy or data protection, and the right to intellectual property. While several examples are given as to how these rights may be implicated in relation to algorithmic speech, ultimately whether a particular right is relevant for a court's balancing exercise is heavily dependent on the factual scenario of the case, as well as the type of algorithm involved.

Finally, other relevant issues surrounding algorithmic speech were found to be important and to have an impact going forward, many of which involve questions of policy and societal values that accompany granting algorithmic speech protection. Among these are the actual form of algorithmic accountability that may potentially be adopted, as well as its intrusiveness and the resulting effects of its implementation, the nature and importance of the service that is making use of the algorithm, the difficulties the border cases may bring, and the impact that protecting algorithmic speech may have on innovation.

The primary research question for this Article was posed as follows: To what extent is algorithmic speech protected under international standards of freedom of expression? The answer here is the same as it is for most legal questions: it depends. The type of algorithm, the nature of the interference, and the factual scenario—including other relevant rights and interests—will affect the extent to

which algorithmic speech is protected. While courts and legislators may have difficulties managing issues relating to algorithmic speech, the current framework for freedom of expression will be adequate, if properly adapted, for some time to come.