DO PLATFORMS KILL?

MICHAL LAVI*

*"So we connect more people[.] That can be bad if they make it negative. Maybe it costs a life by exposing someone to bullies. Maybe somebody dies in a terrorist attack coordinated on our tools. And still we connect people. The ugly truth is that we believe in connecting people so deeply that anything that allows us to connect more people more often is *de facto* good. It is perhaps the only area where the metrics do tell the true story as far as we are concerned."*[1]

Terror kills, inciting words can kill, but what about online platforms? In recent years, social networks have turned into a new arena for incitement. Terror organizations operate active accounts on social networks. They incite, recruit, and plan terror attacks by using online platforms. These activities pose a serious threat to public safety and security.

Online intermediaries, such as Facebook, Twitter, YouTube, and others provide online platforms that make it easier for terrorists to meet and proliferate in ways that were not dreamed of before. Thus, terrorists are able to cluster, exchange ideas, and promote extremism and polarization. In such an environment, do platforms that host in-

---

* Ph.D.(Law); Postdoctoral Fellow, University of Haifa, Faculty of Law; Cyberlaw Fellow, Federmann Cyber Center Hebrew University; Cheshin Fellow, Hebrew University, Faculty of Law, 2018. The author thanks Jonathan Lewy, Michal Shur-Ofry, and Dorit Zoldan-Zohar. Special thanks are due to Nicole Baade, Editor-in-Chief, Jacob Thackston, Joey Montgomery, Alan Chan, William Flanagan, Jonathan DeWitt, Jacob Richards, Adam Sharf, Bryan Poellot, Christian Hecht, Hunter Pearl, Eli Nachmany, Kevin Lie, Jason Altabet, Issac Sommers, Nick Cordova, Wentao Zhai, Payton Alexander, Aaron Ward, Jessica Tong, Alex Riddle, Aaron Henricks, Oliver Roberts, John Mitzel, Bryan Sohn, Brian Kulp, Jasjaap Sidhu, Anna Lukina, Alex Cave, Doug Stephens IV, John Ketcham, and Aaron Gyde on the *Harvard Journal of Law & Public Policy* staff for helpful comments, suggestions, and outstanding editorial work.

1. Ryan Mac, Charlie Warzel & Alex Kantrowitz, Growth At Any Cost: Top Facebook Executive Defended Data Collection In 2016 Memo—And Warned That Facebook Could Get People Killed, BUZZFEED NEWS (Mar. 29, 2018, 6:36 PM), https://www.buzzfeednews.com/article/ryanmac/growth-at-any-cost-top-facebook-executive-defended-data [https://perma.cc/LQ4G-MXFA] (quoting memo by Andrew Bosworth, Vice President of Facebook).

citing content bear any liability? What about intermediaries operating internet platforms that direct extremist and unlawful content at susceptible users, who, in turn, engage in terrorist activities? Should intermediaries bear civil liability for algorithm-based recommendations on content, connections, and advertisements? Should algorithmic targeting enjoy the same protections as traditional speech?

This Article analyzes intermediaries' civil liability for terror attacks under the anti-terror statutes and other doctrines in tort law. It aims to contribute to the literature in several ways. First, it outlines the way intermediaries aid terrorist activities either willingly or unwittingly. By identifying the role online intermediaries play in terrorist activities, one may lay down the first step towards creating a legal policy that would mitigate the harm caused by terrorists' incitement over the internet. Second, this Article outlines a minimum standard of civil liability that should be imposed on intermediaries for speech made by terrorists on their platforms. Third, it highlights the contradictions between intermediaries' policies regarding harmful content and the technologies that create personalized experiences for users, which can sometimes recommend unlawful content and connections.

This Article proposes the imposition of a duty on intermediaries that would incentivize them to avoid the creation of unreasonable risks caused by personalized algorithmic targeting of unlawful messages. This goal can be achieved by implementing effective measures at the design stage of a platform's algorithmic code.

Subsequently, this Article proposes remedies and sanctions under tort, criminal, and civil law while balancing freedom of speech, efficiency, and the promotion of innovation. The Article concludes with a discussion of complementary approaches that intermediaries may take for voluntarily mitigating terrorists' harm.

INTRODUCTION

On June 12, 2016, Omar Mateen committed an attack at an LGBT nightclub in Orlando, Florida.[2] Forty-nine people died along with Mateen.[3] Fifty-three others were injured.[4] On the day of the attack, Mateen posted on Facebook his allegiance to ISIS and demanded that the United States and Russia "stop bombing the Islamic state [sic]."[5] He also warned that further attacks would come: "The real muslims [sic] will never accept the filthy ways of the west . . . . In the next few days, you will see attacks from the Islamic state [sic] in the usa [sic]."[6]

---

2. Crosby v. Twitter, Inc., 303 F. Supp. 3d 564, 567 (E.D. Mich. 2018), *aff'd*, 921 F.3d 61 (6th Cir. 2019).

3. *Id.*

4. *Id.*

5. David Smith & Spencer Ackerman, *Orlando gunman searched for Facebook reaction during Pulse nightclub attack*, GUARDIAN (June 16, 2016, 1:02 PM), https://www.theguardian.com/us-news/2016/jun/16/orlando-attack-facebook-post-pulse-nightclub-shooting [https://perma.cc/9L5T-PJ2F] (internal quotation marks omitted).

6. *Id.*

ISIS claimed responsibility for the shootings shortly thereafter.[7] According to a complaint filed by the victims of the attack, "FBI analysts found that Mateen watched online jihadist sermons since at least 2012 and more recently had downloaded jihadist material to his laptop . . . ."[8] Though there was no evidence he had ever been in contact with ISIS directly, it appears that ISIS was able, at least in part, to radicalize Mateen through the internet.[9]

In November 2015, Anwar Abu Zaid, Jordanian police captain, "shot and killed two government contractors on an American base in Jordan."[10] According to Abu Zaid's brother, Abu Zaid turned to terrorism after watching a video ISIS posted in February 2015, which showed the execution of a Jordanian pilot.[11] ISIS claimed responsibility for the attack.[12] A few months earlier, on December 2, 2015, Syed Rizwan Farook and Tashfeen Malik, a married couple, fired more than 100 rounds into a staff meeting of the environmental health department in San Bernardino, California, murdering fourteen and injuring twenty-two.[13] During the shooting, Tashfeen Malik pledged her loyalty on Facebook to Abu Bakr al-Baghdadi, the leader of ISIS.[14] A couple of days later, ISIS endorsed their acts of terrorism.[15] The FBI investigation of this terror attack revealed that Farook and

---

7. Complaint at 43, *Crosby*, 303 F. Supp. 3d 564 (No. 16-14406) [hereinafter *Crosby* Complaint]. Please note that there are no pleaded facts that Mateen carried out the act under express directions from ISIS. *See id.* at 40–46.

8. *Id.* at 44–45 ("The FBI believes that the Orlando shooter Omar Mateen was self-radicalized on the Internet and social media.").

9. *See id.* at 44; Ed Pilkington & Dan Roberts, *FBI and Obama confirm Omar Mateen was radicalized on the internet*, GUARDIAN (June 14, 2016, 2:06 PM), https://www.theguardian.com/us-news/2016/jun/13/pulse-nightclub-attack-shooter-radicalized-internet-orlando [https://perma.cc/AC74-Y9US].

10. Jaime M. Freilich, Note, *Section 230's Liability Shield in the Age of Online Terrorist Recruitment*, 83 BROOK. L. REV. 675, 676–77, 685 (2018).

11. Fields v. Twitter, Inc., 200 F. Supp. 3d 964, 967 (N.D. Cal. 2016).

12. Fields v. Twitter, Inc., 881 F.3d 739, 742 (9th Cir. 2018).

13. Clayborn. v. Twitter, Inc., Nos. 17-cv-06894-LB & 18-cv-00543-LB, 2018 WL 6839754, at *1, *3 (N.D. Cal. Dec. 31, 2018).

14. *Id.* at *2. Al-Baghdadi died recently during a raid conducted by U.S. military forces in northwest Syria. Eliza Mackintosh, *ISIS leader Abu Bakr al-Baghdadi is dead. Here are 6 things you need to know*, CNN (Oct. 29, 2019, 12:30 PM), https://edition.cnn.com/2019/10/28/middleeast/baghdadi-isis-leader-dead-explainer-intl/index.html [https://perma.cc/AN94-WQ7R].

15. *Clayborn*, 2018 WL 6839754, at *2.

Malik were radicalized by social media platforms several years before the attack.[16]

Social media platforms allow anyone to post content online. In recent years, social media has become a common venue for the dissemination of terrorist propaganda, as well as the radicalization, glorification, and incitement of terrorism.[17] Terror organizations, such as ISIS, Al-Qaeda, Hamas, and white supremacist terrorists,[18] exploit social media to solicit funds for

16. *Id.* at *3.

17. Susan Klein & Crystal Flinn, *Social Media Compliance Programs and the War Against Terrorism*, 8 HARV. NAT'L SECURITY J. 53, 65 & n.55 (2017) (referring to the statement of Nicholas J. Rasmussen, Director of the National Counterterrorism Center: "This online environment is likely to play a critical role in the foreseeable future in radicalizing and mobilizing [Homegrown Violent Extremists] towards violence."(alteration in original) (internal quotation marks omitted)); *see also* Alexander Tsesis, *Social Media Accountability for Terrorist Propaganda*, 86 FORDHAM L. REV. 605, 608 (2017).

18. White supremacist terrorists commit mass murder hate attacks against Muslims, immigrants, Jews, and other groups that they perceive as a threat to their race. As the recent white supremacist terror attacks in Pittsburgh, Pennsylvania; Christchurch, New Zealand; San Diego, California; and El Paso, Texas demonstrate, these terrorists are no less deadly. *See, e.g.*, Kristen Gelineau & Jon Gambrell, *New Zealand mosque shooter is a white nationalist who hates immigrants, documents and video reveal*, CHICAGO TRIBUNE (Mar. 15, 2019, 8:45 PM), https://www.chicagotribune.com/nation-world/ct-mosque-killer-white-supremacy-20190315-story.html [https://perma.cc/T4RK-Q9KS]; Campbell Robertson, Christopher Mele & Sabrina Tavernise, *11 Killed in Synagogue Massacre; Suspect Charged With 29 Counts*, N.Y. TIMES (Oct. 27, 2018), https://nyti.ms/2JlIq5U [https://perma.cc/8VLJ-884C]. Brenton Tarrant, the alleged shooter in the Christchurch mosque attacks, praised prominent Australian far-right nationalist Blair Cottrell on Facebook and referred to him as "Emperor." Alex Mann et al., *Christchurch shooting accused Brenton Tarrant supports Australian far-right figure Blair Cottrell*, ABC NEWS (Mar. 23, 2019, 4:21 AM), https://www.abc.net.au/news/2019-03-23/christchurch-shooting-accused-praised-blair-cottrell/10930632 [https://perma.cc/3K3D-9TVD]. Before committing a deadly attack, John Earnest published a racist open letter on an online forum 8chan, a racist alt-right message board. Michael McGowan, *San Diego shooting suspect posted 'open letter' online*, GUARDIAN (Apr. 28, 2019, 3:25 PM), https://www.theguardian.com/us-news/2019/apr/28/john-earnest-san-diego-shooting-suspect-posted-open-letter-online [https://perma.cc/7S3A-VJVC]. Earnest was inspired by the shooter in New Zealand. *Id.* Tarrant and Earnest are not the only terrorists who were radicalized and posted on 8chan. *See, e.g.*, Robert Evans, *Ignore The Poway Synagogue Shooter's Manifesto: Pay Attention To 8chan's /pol/ Board*, BELLINGCAT (Apr. 28, 2019), https://www.bellingcat.com/news/americas/2019/04/28/ignore-the-poway-synagogue-shooters-manifesto-pay-attention-to-8chans-pol-board/ [https://perma.cc/J6VZ-ZVM7]; Brianna Sacks & Adolfo Flores, *The Suspected El Paso Terrorist Said He Was Motivated By A Hatred Of Immigrants*, BUZZFEED NEWS (Aug. 4, 2019, 7:49 PM), https://www.buzzfeednews.com/article/briannasacks/el-paso-shooting-suspect-immigrants-hate-manifesto [https://perma.cc/K72Z-6CCJ].

their activities.[19] They upload photos and videos of terror attacks in real time, including livestreaming deadly terror attacks that gamify massacring,[20] which reach sympathizers and send propaganda to draw in people who are inclined to radicalization.[21] The recent Walmart terror attack in El Paso, Texas serves as a good example. The killer, Patrick Crusius, announced the start of his rampage on 8chan's board through a post that included a four-page manifesto.[22] The manifesto and posts on 8chan demonstrate Crusius's radicalization and turn towards white supremacy.[23] Based on a review and analysis of 8chan posts, Bellingcat, an investigative journalism website, concluded that an earlier manifesto of the Christchurch's shooter in New Zealand and the video of his attack, likely had a profound influence on Crusius.[24]

Social media allows terrorist groups to reach potential recruits[25] and inspire loners to commit attacks.[26] This use of social

---

19. For example, the Twitter account @Jahd_bmalk solicited donations for weapons with the slogan "Participate in Jihad with your Money." Corrected Complaint at 21, *Clayborn*, No. 17-cv-06894-LB [hereinafter *Clayborn* Corrected Complaint].

20. *See, e.g.*, Freilich, *supra* note 10, at 693 n.140; Lizzie Dearden, *Germany synagogue shooting: Suspect 'broadcast attack livestream on Twitch' and ranted about Holocaust, Jews and immigration*, INDEPENDENT (Oct. 9, 2019, 6:12 PM), https://www.independent.co.uk/news/world/europe/germany-shooting-synagogue-attack-latest-twitch-livestream-gunman-holocaust-jews-a9149381.html [https://perma.cc/S46Y-WBQ5]; Robert Evans, *The El Paso Shooting and the Gamification of Terror*, BELLINGCAT (Aug, 4, 2019), https://www.bellingcat.com/news/americas/2019/08/04/the-el-paso-shooting-and-the-gamification-of-terror/ [https://perma.cc/C94C-SDBG] ("Brenton Tarrant livestreamed his massacre from a helmet cam in a way that made the shooting look almost exactly like a First Person Shooter video game. This was a conscious choice, as was his decision to pick a sound-track for the spree that would entertain and inspire his viewers."); Meagan Flynn, *No one who watched New Zealand shooter's video live reported it to Facebook, company says*, WASH. POST (Mar. 19, 2019, 5:04 AM), https://www.washingtonpost.com/nation/2019/03/19/new-zealand-mosque-shooters-facebook-live-stream-was-viewed-thousands-times-before-being-removed/ [https://perma.cc/M4ZK-8933].

21. *See, e.g.*, J.M. Berger, *How terrorists recruit online (and how to stop it)*, BROOKINGS (Nov. 9, 2015), https://www.brookings.edu/blog/markaz/2015/11/09/how-terrorists-recruit-online-and-how-to-stop-it/ [https://perma.cc/K7CX-L68H].

22. Tim Arrango, Nicholas Bogel-Burroughs & Katie Benner, *Minutes Before El Paso Killing, Hate-Filled Manifesto Appears Online*, N.Y. TIMES (Aug. 3, 2019), https://nyti.ms/2OEIGDs [https://perma.cc/VH2Y-AZQG].

23. *See id.*

24. *See* Evans, *supra* note 20.

25. Tsesis, *supra* note 17, at 617 ("The French interior minister recently asserted that 90 percent of people who are recruited to terrorism are indoctrinated by internet content."); *see also* Paul Gill et al., *Terrorist Use of the Internet by the Numbers:*

media allows terrorists to shock, threaten, communicate ideology, and affect the conduct of millions of viewers. It opens the gateway to violent extremism and incites individuals and groups to commit violence and hate crimes,[27] even if they are not part of a traditional terrorist cell. Incitement on social media has consequences in the physical world, as terrorists increasingly rely on social media to plan and execute attacks.[28] Social media platforms allow terror organizations to operate accounts in their own names, although many of them have been officially dubbed as terrorists.[29]

Clustering like-minded people online accelerates interpersonal dynamics of incitement across the network and enhances polarization and extremism. It increases the likelihood for more people to be engaged in terror attacks.[30] Yet, online intermediaries fail to remove inciting posts in many cases and fail to keep inciting content down.[31]

Quantifying Behaviors, Patterns, and Processes, 16 CRIMINOLOGY & PUB. POL'Y 99, 107–09 (2017).

26. Jade Hutchinson, *Far-Right Terrorism: The Christchurch Attack and Potential Implications on the Asia Pacific Landscape*, 11 COUNTER TERRORIST TRENDS & ANALYSES, June 2019, at 19, 19 ("[I]t is found that the assailant's relationship with the far-right virtual community and attitude towards venerating the online sub-culture, along with his proficiency with Internet technology and mass-violence weaponry, is significant for far-right terrorist behaviour in the Asia Pacific region . . . ."); Martin Rudner, "*Electronic* Jihad*": The Internet as Al Qaeda's Catalyst for Global Terror*, 40 STUD. CONFLICT & TERRORISM 10, 15 (2017) ("The Internet has been noticeably instrumental for Al Qaeda in its ongoing efforts to foster locally homegrown terrorist activities directed against British, European, and North American targets.").

27. Thane Rosenbaum, *The Internet as Marketplace of Madness—And A Terrorist's Best Friend*, 86 FORDHAM L. REV. 591, 594 (2017) ("Without the internet, terrorist cells had as much visibility as actual microorganisms. Without cyberspace, learning how to make a bomb from household detergents had the same degree of difficulty as traveling to outer space. . . . YouTube turned them into genocidal reality TV stars. It was the Wild West of terrorism . . . .").

28. Zachary Leibowitz, Note, *Terror on Your Timeline: Criminalizing Terrorist Incitement on Social Media through Doctrinal Shift*, 86 FORDHAM L. REV. 795, 797 (2017).

29. Cohen v. Facebook, Inc., 252 F. Supp. 3d 140, 147 (E.D.N.Y. 2017).

30. *See* CASS R. SUNSTEIN: #REPUBLIC: DIVIDED DEMOCRACY IN THE AGE OF SOCIAL MEDIA 238, 241 (2017).

31. *See, e.g.*, Yitzhak Benhorin, *20,000 Israelis sue Facebook*, YNET NEWS (Oct. 27, 2015, 8:47 PM), https://www.ynetnews.com/articles/0,7340,L-4716980,00.html [https://perma.cc/EN7C-RRPP]; *see also* Freilich, *supra* note 10, at 676 ("[S]ocial media companies have generally taken a 'laissez-faire approach' to preventing terrorists from using their platforms to promote their illegal agendas . . . ."); Klein & Flinn, *supra* note 17, at 71–72 ("Continued failure to address terror activity online will undoubtedly lead to increased vigilante justice by independent hack-

In addition to terrorists' "bottom-up" social dynamics on social networks, intermediaries enable terrorists' activities from the "top down." Recently, at the Anti-Defamation League, actor and comedian Sacha Baron Cohen criticized social media companies, aptly describing Facebook as "the greatest propaganda machine in history."[32]

Intermediaries profit from terrorists, as they strategically target specific organic content and advertisements based on viewers and content.[33] Some intermediaries share revenues earned from targeted ads with those who posted the content, or with webpage owners.[34] The posters and owners might be terror organizations, and as a result, the shared revenues could fund terrorist activities.[35]

Moreover, in their quest to enhance profits from content and advertisement, intermediaries personalize content by automatic algorithms that recommend additional content to users.[36] These recommendation systems do not "know" what a particular user might prefer, but rather draw conclusions based on past interactions of similar users.[37] Thus, they direct users to new content, which might be terrorist oriented. Intermediaries use these algorithms to connect users with others who might have

---

ers, pulling control and ability to monitor from the government and creating uncertainty in the current methodology used to combat terrorism online.").

32. *See* Sacha Baron Cohen, *Read Sacha Baron Cohen's scathing attack on Facebook in full: 'greatest propaganda machine in history'*, GUARDIAN (Nov. 22, 2019, 1:10 PM), https://www.theguardian.com/technology/2019/nov/22/sacha-baron-cohen-facebook-propaganda [https://perma.cc/XC3H-J2ZW].

33. David Patrikarakos, *Social Media Networks Are the Handmaiden to Dangerous Propaganda*, TIME (Nov. 2, 2017), https://time.com/5008076/nyc-terror-attack-isis-facebook-russia/ [https://perma.cc/B6BS-H2TZ].

34. *See e.g.*, *YouTube channel monetization policies*, YOUTUBE (Jan. 2019), https://support.google.com/youtube/answer/1311392?hl=en [https://perma.cc/GQ49-HHAP].

35. Freilich, *supra* note 10, at 678 ("Though Twitter, Facebook, and Google may not be giving money to terrorist groups, per se, they are giving terrorist groups a platform to spread their violent rhetoric and they are profiting from those groups' presence on their websites.").

36. Derek O'Callaghan et al., *Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems*, 33 SOC. SCI. COMPUTER REV. 459, 460 (2015); *see also* Kevin Roose, *The Making of a YouTube Radical*, N.Y. TIMES (June 8, 2019), https://nyti.ms/2wygCsx [https://perma.cc/ZBV2-XS77]; Joan E. Solsman, *YouTube's AI is the puppet master over most of what you watch*, CNET (Jan. 10, 2018, 10:05 AM), https://www.cnet.com/news/youtube-ces-2018-neal-mohan [https://perma.cc/7XTS-7RA9].

37. VIKTOR MAYER-SCHÖNBERGER & THOMAS RAMGE, REINVENTING CAPITALISM IN THE AGE OF BIG DATA 78 (2018). "These systems don't understand the data in any human sense; they only identify the patterns they are 'seeing' . . . ." *See id.*

shared interests, even if the results of these match-ups go against the websites' content moderation policies.[38] This practice can play a vital role in spreading inciting content to those users most susceptible to that incitement.

The practice of targeted recommendation by the "AI propaganda machine"[39] may encourage susceptible social network members to consume extreme and even inciting content.[40] Targeted algorithmic-based recommendations increase the likelihood of influencing users because they seek the recommended content and are more susceptible to it.[41] Inciting content can thus radicalize susceptible social network users, and they are more likely to disseminate the inciting content and even act upon it. This may result in more victims of terror.

Terror victims and their families have brought suits against intermediaries, arguing that the offensive content, the practice of revenue sharing with terror organizations, and the personalization of recommendations to susceptible social network members materially supports terrorism in violation of federal antiterrorism laws.[42] In other words, the plaintiffs asserted that intermediaries were responsible for the physical harm and death caused by terrorists.

Should the law impose civil liability on intermediaries for terror attacks and allow victims to get redress? And if so, what should be the appropriate scope of intermediaries' civil liability

---

38. For example, an intermediary can block specific types of content and simultaneously recommend them by using automatic algorithms. *See* Ysabel Gerrard, *Beyond the hashtag: Circumventing content moderation on social media*, 20 NEW MEDIA & SOC'Y 4492, 4505 (2018). *See generally* Karl Manheim & Lyric Kaplan, *Artificial Intelligence: Risks to Privacy and Democracy*, 21 YALE J.L. & TECH 106 (2019).

39. Berit Anderson & Brett Horvath, *The Rise of the Weaponized AI Propaganda Machine*, MEDIUM (Feb. 12, 2017), https://medium.com/join-scout/the-rise-of-the-weaponized-ai-propaganda-machine-86dac61668b [https://perma.cc/7AQF-SKQK]; *see also* MILES BRUNDAGE ET AL., THE MALICIOUS USE OF ARTIFICIAL INTELLIGENCE: FORECASTING, PREVENTION, AND MITIGATION 3–11, https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf [https://perma.cc/WUY6-MNJE].

40. *See* O'Callaghan et al., *supra* note 36, at 460; Zeynep Tufekci, Opinion, *YouTube, the Great Radicalizer*, N.Y. TIMES (Mar. 10, 2018), https://nyti.ms/2GeTMa6 [https://perma.cc/E53F-LTQB].

41. *See* MAX TEGMARK, LIFE 3.0: BEING HUMAN IN THE AGE OF ARTIFICIAL INTELLIGENCE 18 (2017) (describing "'persuasion sequences' of videos where insight from each one would both update someone's views and motivate them to watch another video about a related topic where they were likely to be further convinced").

42. *See, e.g.*, Crosby v. Twitter, Inc., 303 F. Supp. 3d 564, 567–68 (E.D. Mich. 2018); Cohen v. Facebook, Inc., 252 F. Supp. 3d 140, 147 (E.D.N.Y. 2017).

and legal duty of care? This Article answers these questions and others. The Article defines terrorism as "the deliberate killing of innocent people, at random, in order to spread fear through a whole population and force the hand of its political leaders."[43] It explores the question of intermediaries' liability for incitement to terrorism on social media websites[44] and focuses on online social networks in particular.[45]

Part I of the Article focuses on the evolution of modern terrorism in the wake of social networks. It describes the influence of terror organization on social dynamics within social networks, which enhances inciting speech that can push participants to commit terror attacks. Part II outlines the different roles intermediaries take in facilitating networks that promote terrorist attacks. Part III explores the civil liability of intermediaries under the federal antiterrorism laws and section 230 of the Communications Decency Act.[46] Following this analysis, this Part deals with normative considerations for imposing liability on intermediaries. Part IV discusses the possibility of imposing liability on online intermediaries for material support of terrorist activities. It proposes a minimum standard for mandatory removal of unlawful content. It also argues that social media platforms can no longer hide behind the notion that they are neutral platforms when their moderation and algorithmic recommendation systems determine what content is seen and heard.[47]

---

43. *See* Michael Walzer, *Five Questions About Terrorism*, DISSENT MAG. (2002), https://www.dissentmagazine.org/article/five-questions-about-terrorism [https://perma.cc/9785-S64V].

44. *See* Jan H. Kietzmann et al., *Social media? Get serious! Understanding the functional building blocks of social media*, 54 BUS. HORIZONS 241, 241 (2011) ("Social media employ mobile and web-based technologies to create highly interactive platforms via which individuals and communities share, co-create, discuss, and modify user-generated content.").

45. *See* danah m. boyd & Nicole B. Ellison, *Social Network Sites: Definition, History, and Scholarship*, 13 J. COMPUTER-MEDIATED COMM. 210, 211 (2008) (defining social network sites as "web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system").

46. 47 U.S.C. § 230 (2018).

47. *See* TARLETON GILLESPIE, CUSTODIANS OF THE INTERNET: PLATFORMS, CONTENT MODERATION AND THE HIDDEN DECISIONS THAT SHAPE SOCIAL MEDIA 24–45 (2018); Danielle Keats Citron, *Section 230's Challenge to Civil Rights and Civil Liberties*, KNIGHT FIRST AMENDMENT INST. COLUM. U. (Apr. 6, 2018), https://

This Article proposes imposing duties on intermediaries that would disincentivize them from taking unreasonable risks and manipulating users by targeting susceptible users with content that radicalizes and incites them to terror. These duties focus on the design stage of the platform, thus creating a regime of "safety by design." This Article also proposes remedies and sanctions under the loss of chance doctrine in tort, criminal, and civil law. In doing so, it accounts for freedom of speech, economic efficiency, and innovation promotion. The Article concludes with complementary tools that intermediaries may voluntarily use to mitigate the harm caused by terrorists' speech.

I. THE EVOLUTION OF NETWORKED TERROR

A. From Localities to Online Social Networks

"Social networks seem to organize social life today."[48] These sets of relationships[49] spread happiness, generosity, and love. They are always there, exerting dramatic influence over choices, actions, thoughts, feelings, and even desires. Social networks may affect the full spectrum of human experience.

Networks have always been the leading force behind terror, even before the internet age. "Social dynamics—not poverty, poor education and disadvantage"—have played and continue to play a central role in the development and diffusion of terrorism.[50] How did terrorists gather before the advent of social media? Where did they meet? In his book, *Understanding Terror*, Marc Sageman, a forensic psychiatrist, former CIA agent, and government counterterrorism consultant, tries to answer these questions.[51] Based on the collection and analysis of data on 400 Islamic terrorists who lived during the 1990s, he demonstrates that many terrorists had families and distinguished jobs.[52] Some were not very religious at the time they joined the Salafi

knightcolumbia.org/content/section-230s-challenge-civil-rights-and-civil-liberties [https://perma.cc/B9BQ-YMAJ].

48. Michal Lavi, *Content Providers' Secondary Liability: A Social Network Perspective*, 26 FORDHAM INTELL. PROP. MEDIA & ENT. L.J. 855, 889 (2016).

49. CHARLES KADUSHIN, UNDERSTANDING SOCIAL NETWORKS: THEORIES, CONCEPTS AND FINDINGS 14 (2012) (explaining that social networks are sets of relationships).

50. SUNSTEIN, *supra* note 30, at 234.

51. MARC SAGEMAN, UNDERSTANDING TERROR NETWORKS 61–98 (2004).

52. *Id.* at 78–80.

jihad, the violent, revivalist social movement including al Qaeda.[53] Seventy percent joined the jihad while they were living away from their country of origin.[54] They sometimes met each other in mosques, not necessarily for religious reasons, but rather to seek friends with similar cultural backgrounds.[55] Some who met at mosque also moved into apartments together and developed a microculture.[56] Their life developed a group dynamic that ultimately transformed them into terrorists.[57] They were not recruited for terror missions but rather volunteered to act.[58] The network was self-organized from the bottom up and the dynamics within it enforced the motivation of the members of the group to engage in terror.[59] The network grew as it gathered more members, who met each other in person.[60] Yet, before the age of social media, the possibility to engage with like-minded people anytime and anywhere was limited, thus reducing the scale of polarization and extremism. Technology and new media weaponized terrorism, and that is what made "terrorism and the internet such a toxic, incitement-spiked brew."[61]

B. Terror-Networks.Com

Networks have always existed, but online networks operate in a different environment. The internet revolution, mobile phones, and social networks enhanced the ability of users to stay in touch with one another constantly and immediately.[62] This revolution afforded new opportunities to form social ties, share ideas, form communities, and engage in diverse social dynamics anywhere, anytime.[63] Technology creates different

53. *Id.* at 61–62, 76–77, 97.
54. *Id.* at 92.
55. *Id.* at 96, 143.
56. *Id.* at 101.
57. *Id.* at 115.
58. *Id.* at 110, 122.
59. *Id.* at 110–12.
60. *See id.* at 99–111.
61. Rosenbaum, *supra* note 27, at 600.
62. *See* NICHOLAS A. CHRISTAKIS & JAMES H. FOWLER, CONNECTED: THE SURPRISING POWER OF OUR SOCIAL NETWORKS AND HOW THEY SHAPE OUR LIVES 275 (2009).
63. *See* GILLESPIE, *supra* note 47, at 5 ("Social media platforms put more people in direct contact with one another, afford them new opportunities to speak and interact with a wider range of people, and organize them into networked publics.").

tools that influence beliefs, preferences, and capabilities in society.[64] "The medium matters because it shapes, structures, and controls the scale, scope, reach, pace, and patterns of human communications . . . ."[65]

Before the terror attacks on September 11, 2001, terror organizations radicalized through face-to-face interactions. These interactions have been "replaced by online radicalization."[66] The internet has made it easier than ever to overcome geographical barriers and establish contacts among terrorist groups that are far apart in the physical world.[67]

Social media now enables terrorist organizations to expand and amplify their presence on the world stage.[68] Online platforms provide terrorists "the means to collaborate, share membership lists, recruit new members, and advise each other."[69] As research demonstrates, social media allows self-organized groups that have probably never met in person before to increase their numbers and inspire others to carry out attacks.[70] Today, there is no doubt that communication by online networks dramatically influences "how the message of extremism is conveyed."[71] Social media takes terrorism to a different scale,

---

64. BRETT FRISCHMANN & EVAN SELINGER, RE-ENGINEERING HUMANITY 47, 106 (2018).

65. *Id.* at 107; *see also* Michael J. Sherman, Brandenburg *v. Twitter*, 28 GEO. MASON U. C.R. L.J. 127, 131 (2018) ("[T]here are at least two significant characteristics that make online recruitment of terrorists different: the ability to reach mass audiences is vastly greater than it was a generation or two ago, and there may be greater violence tied to the speech in question.").

66. *Violent Islamist Extremism: Hearing Before the S. Comm. on Homeland Sec. & Governmental Affairs*, 110th Cong. 473 (2007) (written statement of Dr. Marc Sageman, Principal, Sageman Consulting, LLC) [hereinafter *Violent Islamist Extremism Hearing*].

67. *See* Sherman, *supra* note 65, at 131–32.

68. Lauren C. O'Leary, Note, *Targeting Detached Corporate Intermediaries in the Terrorist Supply Chain: Dial 2339/13224 for Assistance*, 103 VA. L. REV. 525, 561 (2017).

69. Alexander Tsesis, *Terrorist Speech on Social Media*, 70 VAND. L. REV. 651, 654 (2017).

70. *See Tracking, analyzing how ISIS recruits through social media*, HOMELAND SECURITY NEWS WIRE (June 20, 2016), http://www.homelandsecuritynewswire.com/dr20160620-tracking-analyzing-how-isis-recruits-through-social-media [https://perma.cc/AQ73-MCZ6].

71. Amos N. Guiora, *Inciting Terrorism on the Internet: The Limits of Tolerating Intolerance*, *in* INCITEMENT TO TERRORISM 137, 138 (Anne F. Bayefsky & Laurie R. Blank eds., 2018); *see also* Julie E. Cohen, *The emergent limbic media system*, *in* LIFE AND THE LAW IN THE ERA OF DATA-DRIVEN AGENCY 60, 74 (Mireille Hildebrandt & Kieron O'Hara eds., 2020) ("The increasingly unreasoning and often vicious char-

as demonstrated in recent terror attacks in the United States, France, New Zealand, Israel and many other countries.[72]

New patterns of social connections unique to online culture play a role in spreading modern terrorism. Terror organizations create social structures that regenerate themselves and do not depend on a single leader. Online activists can connect with each other despite being scattered around the globe.[73] As a new study confirms, activists can start spreading their word on ideological social media websites, such as the far alt right websites Gab[74] and 8chan.[75] As more people follow others and repeat their inciting messages, they allow such content to penetrate mainstream social media, gain influence, incite more people, and shape pathways to violence on larger, general platforms.[76]

---

acter of interaction in online, platform-based digital environments complicates accounts of the democratizing potential of information networks."); Lyrissa Barnett Lidsky, *Incendiary Speech and Social Media*, 44 TEX. TECH L. REV. 147, 149 (2011) (explaining that the anonymity of inciting communication online "fosters a sense of disinhibition in those contemplating violence").

72. *See, e.g.*, Crosby v. Twitter, Inc., 303 F. Supp. 3d 564, 567–68 (E.D. Mich. 2018) (describing the terror attack at an LGBT nightclub in Orlando); Gonzalez v. Google, Inc., 282 F. Supp. 3d 1150, 1154 (N.D. Cal. 2017) (describing the ISIS attack in Paris, France); Mann, *supra* note 18; Siobhán O'Grady, *Families of Americans Killed in Israel Sue Facebook For $1 Billion*, FP (July 11, 2016, 4:37 PM), https://foreignpolicy.com/2016/07/11/families-of-americans-killed-in-israel-sue-facebook-for-1-billion/ [https://perma.cc/X3AQ-E38R].

73. SUNSTEIN, *supra* note 30, at 242.

74. *See* Emma Grey Ellis, *Gab, the Alt-Right's Very Own Twitter, Is the Ultimate Filter Bubble*, WIRED (Sept. 14, 2016, 7:00 AM), https://www.wired.com/2016/09/gab-alt-rights-twitter-ultimate-filter-bubble/ [https://perma.cc/VWR3-WX5M]; David Gilbert, *Here's How Big Far Right Social Network Gab Has Actually Gotten*, VICE NEWS (Aug. 16, 2019, 2:35 PM), https://www.vice.com/en_us/article/pa7dwg/heres-how-big-far-right-social-network-gab-has-actually-gotten [https://perma.cc/MTZ3-L5XT].

75. 8chan is an image board site popular with extremists. *See* Ian Sherr & Daniel Van Boom, *8chan is struggling to stay online after El Paso massacre*, CNET (Aug. 7, 2019, 6:21 AM), https://www.cnet.com/news/8chan-is-struggling-to-stay-online-in-wake-of-el-paso-massacre/ [https://perma.cc/EXM9-MV5B].

76. *See* NETWORK CONTAGION RESEARCH INST. & ADL'S CTR. ON EXTREMISM, *Gab and 8chan: Home to Terrorist Plots Hiding in Plain Sight*, ADL (2019), https://www.adl.org/resources/reports/gab-and-8chan-home-to-terrorist-plots-hiding-in-plain-sight [https://perma.cc/77X6-FDBM]. This research shows that online propaganda can inspire terror, and violent terror attacks can perpetuate online propaganda. *Id.* It is in line with known principles of network theory and the concept of "threshold" to explain how ideas spread. *See* Mark Granovetter, *Threshold Models of Collective Behavior*, 83 AM. J. SOC. 1420, 1422 (1978) (explaining that different individuals require different levels of safety for joining an activity, such as entering a riot, and vary in the benefits they derive from the activity).

Terrorist leaders and activists also connect with users, impose psychological pressure on them, and amplify their preexisting inclinations. Consequently, users hear louder echoes of their own voices, and their confirmation bias is amplified as their beliefs are enforced.[77] These users are likely to circulate stories and messages that they agree with and thereby become more extreme.[78] Polarization of groups causes a cascade effect that flares up terrorism, which feeds the dissemination of ideas and attracts more users in social networks.[79] After joining a terror organization like ISIS, new recruits spread propaganda themselves through their social media accounts. A marketplace for extremist ideas becomes "the virtual 'invisible hand' organizing terrorist activities worldwide."[80]

Since 2009, the use of the Internet for terror recruitment and radicalization has increased exponentially.[81] Terrorists make initial contact, profile the potential recruit, and develop a relationship with him online. Afterwards, they isolate him from his community and keep in regular touch with him.[82] Recruitment can focus on unlikely candidates. For example, a 23-year-old Sunday school teacher was recruited via Twitter, email, and Skype.[83] ISIS answered her questions politely while slowly

77. SUNSTEIN, *supra* note 30, at 123.

78. *Id.* at 155 ("[P]eople are biased to like and to publicize opinions and information (real or apparent) that support what they think. Falsehoods spread rapidly, and to the extent that people are reading and speaking to like-minded others, group polarization is inevitable. It is a fact of life in the networked public sphere."); *see also* AN XIAO MINA, MEMES TO MOVEMENTS: HOW THE WORLD'S MOST VIRAL MEDIA IS CHANGING SOCIAL PROTEST AND POWER 125–29 (2019) (explaining that leaders of movements use confirmation bias to increase disinformation).

79. Informational cascades form when individuals follow the statements or actions of predecessors and do not express their opposing opinions because they believe their predecessors are right. CASS R. SUNSTEIN, INFOTOPIA: HOW MANY MINDS PRODUCE KNOWLEDGE 88–90 (2006). As a result, the social network does not obtain important information. *Id.* at 89–90. Reputational cascades form because of social pressures. *Id.* at 91. In these cases, "people think they know what is right, or what is likely to be right, but they nonetheless go along with the crowd in order to maintain the good opinion of others." *Id.*

80. *Violent Islamist Extremism Hearing*, *supra* note 66, at 474.

81. Klein & Flinn, *supra* note 17, at 65 ("Most recently, ISIS has drawn over 20,000 foreign fighters to Syria from more than 90 countries, mainly through cyber contacts.").

82. *See* SUNSTEIN, *supra* note 30, at 242–45.

83. Rukmini Callimachi, *ISIS and the Lonely Young American*, N.Y. TIMES (June 27, 2015), https://nyti.ms/1BX5HoJ [https://perma.cc/V6BY-U9BM].

pushing her towards an extreme worldview.[84] The recruiters advised her to avoid the local mosque that disavowed ISIS by telling her that the government had infiltrated it, adding to her isolation in real life.[85]

Terrorist organizations also use private communication to plan and execute attacks. Turning from public to private communications, such as encrypted messaging, is referred to as "going dark."[86] Yet, terrorists are likely to continue to flourish in open and public platforms because they aim to target the public and impose fear.[87]

Terrorists' use of social media for propaganda and recruitment purposes is only part of the story. Online intermediaries such as Facebook and others not only offer the platforms that facilitate the relationship between self-radicalized cells and transnational community of terror activists, but also have a role in building systems of unforeseen vulnerabilities and enhancing the proliferation of online incitement.

II. SOCIAL MEDIA PLATFORMS AND TERROR: A DESCRIPTIVE ROADMAP

The Director of the FBI has stated with reference to Twitter, "There is a device—almost a devil on their shoulder—all day long, saying: 'Kill. Kill. Kill. Kill.'"[88] Twenty-first-century intermediaries are not mere passive conduits; they take active roles in manipulating users' content. This Part maps intermediaries' role in shaping users experiences in relations to inciting content. It does not, however, advocate the imposition of liability on intermediaries under all circumstances. On the contrary, some of the roles intermediaries take are an inherent part of operating online platforms for legitimate purposes. The control

84. *Id.*

85. *Id.*

86. Klein & Flinn, *supra* note 17, at 66–67. For an expanded discussion, see GABRIEL WEIMANN, GOING DARKER? THE CHALLENGE OF DARK NET TERRORISM (n.d.), https://www.wilsoncenter.org/sites/default/files/media/documents/publication/going_darker_challenge_of_dark_net_terrorism.pdf [https://perma.cc/UKV8-PLPV].

87. *See* GILLESPIE, *supra* note 47, at 55, 171; Klein & Flinn, *supra* note 17, at 68–69.

88. Hamza Shaban, *FBI Director Says Twitter Is A Devil On The Shoulder For Would-Be Terrorists*, BUZZFEED NEWS (July 8, 2015, 5:15 PM) (quoting FBI Director James Comey) (internal quotation marks omitted), https://www.buzzfeednews.com/article/hamzashaban/fbi-doj-evoke-isis-threat-to-justify-encryption-workarounds [https://perma.cc/UG4C-TK2T].

intermediaries have over users' experience can be divided to three types: intermediation, moderation, and algorithmic targeting.

A. Basic Intermediation: Hosting, Providing Communication Tools, and Sharing Revenues

1. Hosting

General purpose social media intermediaries offer platforms for creating content.[89] They utilize technologies and design tools that allow their users to sort through vast amounts of information and share content. Intermediaries allow users to publish and share all kinds of content and encourage ongoing engagement on their platforms.[90]

Terrorist organizations' use of social media platforms and tools is not new. Traditional media has been reporting the use of Twitter, Facebook, and YouTube by terror organizations for years.[91] Testimonies before Congress indicate the widespread use and exploitation of social media platforms and communication services in recruiting members, soliciting funds, and spreading terrorists' propaganda,[92] including livestreaming of terror attacks in real time, leading to visceral reactions from the audience and increasing the likelihood of sharing them.[93] General purpose platforms host a variety of content, only part of which is incitement; but the use of platforms by terrorists is intensifying.

---

89. This Article focuses on general platforms. There are, however, ideological platforms devoted to incitement and hate. These platforms do more than mere hosting, because they create a focal point for hate speech. For more on these platforms, see Michal Lavi, *Evil Nudges*, 21 VAND. J. ENT. & TECH. L. 1 (2018).

90. Jack M. Balkin, *The First Amendment in the Second Gilded Age*, 66 BUFF. L. REV. 979, 997 (2018) ("[B]ecause social media companies encourage as many people as possible to use their sites, the inevitable result is incivility, trolling, and abuse.").

91. *See, e.g.*, Alex Altman, *Why Terrorists Love Twitter*, TIME (Sept. 11, 2014), https://time.com/3319278/isis-isil-twitter/ [https://perma.cc/K4ZC-PGPP]; Marc Santora & Al Baker, *Brooklyn Arrests Highlight Challenges in Fighting of ISIS and 'Known Wolves,'* N.Y. TIMES (Feb. 28, 2015), https://nyti.ms/1JYRBX3 [https://perma.cc/EYF9-RNLR].

92. Tsesis, *supra* note 17, at 617 ("Testimony before Congress in 2015 indicated that ISIS had over 46,000 Twitter accounts and that its followers sent between 90,000 and 200,000 tweets per day.").

93. *See, e.g.*, Evans, *supra* note 20.

2. Providing Communication Tools

In addition to hosting content, social media platforms provide communication tools. These tools improve communication for all users without preferring one type of content to another.[94] Social media users can utilize these tools for any purpose, whether lawful or unlawful.

The tagging options on social networks such as the Twitter hashtag make it easier to aggregate and find relevant content.[95] "Algorithms like Twitter trending catch [hashtags] and highlight them in a section on the site that is visible to many, which in turn drives more attention."[96] These communication tools allow users to promote specific content such as newsworthy scoops. Terrorists use hashtags to make propaganda available to users. For example, ISIS uses inciting hashtags to make it easier for their supporters to cluster together.[97] ISIS tweeted over 14,000 messages threatening Americans under the hashtags #WaronWhites and #AMessagefromISIStoUS, which included photos of U.S Marines hung from bridges in Fallujah.[98] Other posts include threats to kill all Americans.[99] The Hamas and other terrorist organizations exploit hashtags in a similar manner.[100] Moreover, communication tools allow for the spreading of propaganda, and make that propaganda publicly visible. Recruiters communicate by tweeting, retweeting, and using popular hashtags or hashtags related to other trending news stories, such as the World Cup, to communicate inciting material to a wider audience.[101]

---

94. All designs, however, reflect values and are not completely neutral. *See* WOODROW HARTZOG, PRIVACY'S BLUEPRINT: THE BATTLE TO CONTROL THE DESIGN OF NEW TECHNOLOGIES 21–22 (2018).

95. *See* Gerrard, *supra* note 38, at 4493–94 (focusing on the app Tumblr).

96. MINA, *supra* note 78, at 55.

97. Freilich, *supra* note 10, at 676; SUNSTEIN, *supra* note 30, at 242–45.

98. Christopher Bucktin, *ISIS militants send threatening Twitter messages to US warning of retaliation over Iraq air strikes*, MIRROR (Aug. 9, 2014, 6:51 AM), https://www.mirror.co.uk/news/world-news/isis-militants-send-threatening-twitter-4027095 [https://perma.cc/77JC-Z59E].

99. *Id.*

100. *See* Tsesis, *supra* note 17, at 617 (noting hashtags such as "#slaughter of Jews").

101. *See* Michelle Roter, Note, *With Great Power Comes Great Responsibility: Imposing a Duty to Take Down Terrorist Incitement on Social Media*, 45 HOFSTRA L. REV. 1379, 1388 (2017).

3. Sharing Revenues with Users

Intermediaries often share advertisement revenues with users. For example, YouTube allows users to create Google AdSense accounts and monetize those accounts.[102] If there are ads associated with a YouTube video that had been approved by Google, an ad is presented alongside it.[103] YouTube then shares revenues with the poster for each view of the video.[104] The opportunity to share revenues with the intermediary is open to all users whose AdSense account Google has approved.[105] Thus, terrorist organizations and their affiliates can also benefit from monetizing their accounts. Consequently, social media platforms transfer direct payments to terror organizations' affiliates that operate those accounts, and indirectly support terrorist activities.[106]

B. Moderation: Enforcing Policy, Weeding out Terrorist Content and Accounts (Or Neglecting To Do So)

Content moderation promotes adherence to the platforms' terms of use statements, site guidelines, and legal regimes. It is a key part of the production chain of commercial sites and social media platforms.[107] A body of recent scholarship focuses on content moderation and governance. Professor Tarleton Gillespie posits in his book that intermediaries must moderate content; in fact, he demonstrates that their moderation is a fundamental aspect of any platform.[108] Many interviews with moderators show that moderation is needed for a proper operation of the internet,[109] and social media companies cannot deny that moderation is a critical part of their production chain.[110]

---

102. *See YouTube channel monetization policies*, *supra* note 34.

103. *See id.*

104. *See id.*

105. *See Flexible Revenue Sharing*, GOOGLE DEVELOPERS, https://developers.google.com/adsense/host/revenuesharing [https://perma.cc/R5QU-Y5A4] (last visited Mar. 19, 2020).

106. Ronbert H. Schwartz, *Laying the Foundation for Social Media Prosecutions Under 18 U.S.C. § 2339B*, 48 LOY. U. CHI. L.J. 1181, 1189 (2017).

107. SARAH T. ROBERTS, BEHIND THE SCREEN: CONTENT MODERATION IN THE SHADOW OF SOCIAL MEDIA 71 (2019) (describing how different types of low-wage human contractors moderate content).

108. GILLESPIE, *supra* note 47, at 5–6.

109. *See, e.g.*, ROBERTS, *supra* note 107, at 165.

110. *Id.* at 203.

Social media companies often regulate speech in many different ways, using different tools.[111] As Professor Kate Klonick shows, intermediaries already govern speech, enforce their policies and terms of service, and moderate harmful content,[112] even though they are not obligated to do so.[113] They can moderate content before it is published on their sites (ex ante moderation), or after (ex post moderation).[114] Moderation may be reactive, when it is employed upon notices sent to moderators or proactive when moderators seek out published content for removal.[115] It can be done automatically by software or manually by humans.[116] Indeed, intermediaries can and do moderate content.[117] Facebook even has a global escalations team, which removes heinous images and videos from the platform.[118] However, intermediaries' approaches toward moderation are inconsistent within a given platform,[119] and differ among platforms.[120] Despite news reports regarding the use of social media by terrorists, intermediaries' moderation of terrorist content is insufficient, as it continues to spread on Twitter, Facebook, and YouTube. Social media companies are consciously failing

111. GILLESPIE, *supra* note 47, at 1–15.

112. Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598, 1625–30 (2018) (explaining that reasons for moderation are corporate responsibility and economic reasons).

113. *See* 47 U.S.C. § 230(c) (2018) (allowing content providers to enjoy broad immunity).

114. Klonick, *supra* note 112, at 1635.

115. *Id.*

116. *Id.*

117. *See* GILLESPIE, *supra* note 47, at 116 (explaining the labor of moderation by community flaggers, community manager, AI detection tools, crowd workers, and internal teams).

118. Kate Klonick, *Inside the Team at Facebook that Dealt with the Christchurch Shooting*, NEW YORKER (Apr. 25, 2019), https://www.newyorker.com/news/news-desk/inside-the-team-at-facebook-that-dealt-with-the-christchurch-shooting [https://perma.cc/U4FG-2PMA].

119. GILLESPIE, *supra* note 47, at 117 ("Because this work is distributed among different labor forces, because it is unavailable to public and regulatory scrutiny, and because it is performed under high pressure conditions, there is a great deal of room for slippage, distortions, and failure.").

120. *Id.* at 20 ("Platforms vary, in ways that matter both for the influence they can assert over users and for how they should be governed."). Most social media platforms prohibit violence and illegal activity in their policy guidelines, but they differ in definition of these types of content. *Id.* at 54–60.

to combat the use of their websites to promote terrorism and other abuses of the platform.[121]

Intermediaries are inconsistent in removing terrorists' harmful content and hashtags. Twitter used to take a laissez-faire approach to terrorist content and avoided removing it even if it was made aware of the content.[122] Even Facebook, which has a policy against inciting content, does not take a consistent line toward the removal of terrorist content.[123]

Facebook allows access to degrading statements on their platforms despite its own community standards. Although Facebook administrators received several requests to take down a graphic page called "Stab Israelis" and similar inciting pages, it neglected to abide by its own written policy against posting statements favoring brutal attacks, and did not remove these explicit calls for violence.[124] YouTube also allows ISIS accounts and videos on the platform, even though these accounts conflict with its policies.[125] Requests that YouTube voluntarily remove videos of militant terror groups have enjoyed limited success.[126]

---

121. *See* SEC'Y OF STATE FOR THE HOME DEP'T, RADICALISATION: THE COUNTER-NARRATIVE AND IDENTIFYING THE TIPPING POINT 4 (2017), https://www.parliament.uk/documents/commons-committees/home-affairs/Correspondence-17-19/Radicalisation-the-counter-narrative-and-identifying-the-tipping-point-government-response-Eighth-Report-26-17-Cm-9555.pdf [https://perma.cc/5LP6-QPBJ]; *see also* Danielle Keats Citron, *Cyber Mobs, Disinformation, and Death Videos: The Internet as It Is (and as It Should Be)*, 118 MICH L. REV. (forthcoming) (manuscript at 2) (book review) ("Right now, it is cheap and easy to wreak havoc online and for that havoc to go viral. Platforms act rationally . . . when they tolerate abuse that earns them advertising revenue and costs them nothing in legal liability.").

122. *See* Nina I. Brown, *Fight Terror, Not Twitter: Insulating Social Media From Material Support Claims*, 37 LOY. L.A. ENT. L. REV. 1, 10 (2016). Twitter has since taken a more aggressive approach, following the White House's official encouragement of social media platforms to block more terrorists from using their services. *See* Roter, *supra* note 101, at 1391–92, 1398. Twitter's efforts, however, do not satisfy international institutions such as the European Commission, which recently criticized it. *See* Sherman, *supra* note 65, at 147.

123. Roter, *supra* note 101, at 1399.

124. *See* Tsesis, *supra* note 17, at 60. Although Facebook initially refused to eliminate the "Stab Israelis" page, it eventually complied after an Israeli newspaper printed information about the company's intransigence. JNS.org, *Facebook Removes 'Stab Israelis' Page Following Article in Hebrew Press*, ALGEMEINER (Oct. 14, 2015, 2:47 PM), https://www.algemeiner.com/2015/10/14/facebook-removes-stab-israelis-page-following-article-in-hebrew-press/ [https://perma.cc/5DNN-VG4L].

125. Tsesis, *supra* note 17, at 611–12.

126. *Id.*

Many technology experts agree that intermediaries' efforts to proactively detect and remove terrorists' content by utilizing technology also fall short.[127] Although intermediaries use technology to moderate harmful content by preventing upload or re-upload in related contexts,[128] such as copyright[129] and revenge porn,[130] they fail to develop and utilize sufficient technology for addressing terrorist-inciting content.

Moreover, intermediaries can utilize the very same data-driven technology they use to target their users with advertisements and enhance their profits to promote efficient identification and removal of terrorist content,[131] so long as intermediaries have incentives to do so. But, thus far, technological suggestions for moderation of terrorist content have been rejected.[132] There is concern that algorithms will fail to capture context accurately, resulting in both over-removal of content that is not incitement, but lawful information ("false positives"),[133] and under-

127. *See* Nicole Perlroth & Mike Isaac, *Terrorists Mock Bids to End Use of Social Media*, N.Y. TIMES (Dec. 7, 2015), https://nyti.ms/1lKDwSF [https://perma.cc/SR4A-RPSP] (providing the view of Hany Farid, chairperson of the computer science department at Dartmouth College, that the tracking system for child pornography he developed with Microsoft can be applied to terror content). *But see* GILLESPIE, *supra* note 47, at 98–101 ("State-of-the-art detection algorithms have a difficult time discerning offensive content or behavior even when they know precisely what they are looking for . . . .").

128. Klonick, *supra* note 112, at 1635.

129. Maayan Perel & Niva Elkin-Koren, *Accountability in Algorithmic Copyright Enforcement*, 19 STAN. TECH. L. REV. 473 (2016); *see also* John Paul Titlow, *YouTube is using AI to police copyright—to the tune of $2 billion in payouts*, FAST COMPANY (July 13, 2016), https://www.fastcompany.com/4013603/youtube-is-using-ai-to-police-copyright-to-the-tune-of-2-billion-in-payouts [https://perma.cc/8AR8-3LFF]. The EU even enacted a directive that includes filtering requirement. *See* Michelle Kaminsky, *EU's Copyright Directive Passes Despite Widespread Protests—But It's Not Law Yet*, FORBES (Mar. 26, 2019, 1:15 PM), https://www.forbes.com/sites/michellekaminsky/2019/03/26/eus-copyright-directive-passes-despite-widespread-protestsbut-its-not-law-yet/ [https://perma.cc/MQ8U-B2FH].

130. *See* Olivia Solon, *Facebook asks users for nude photos in project to combat 'revenge porn,'* GUARDIAN (Nov. 7, 2017, 5:16 PM), https://www.theguardian.com/technology/2017/nov/07/facebook-revenge-porn-nude-photos [https://perma.cc/P6GC-R8VJ].

131. Tsesis, *supra* note 17, at 611.

132. *See, e.g.*, Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035, 1043–45 (2018).

133. *See* DAPHNE KELLER, DOLPHINS IN THE NET: INTERNET CONTENT FILTERS AND THE ADVOCATE GENERAL'S *GLAWISCHNIG-PIESCZEK V. FACEBOOK IRELAND* OPINION 18 (2019), https://cyberlaw.stanford.edu/files/Dolphins-in-the-Net-AG-Analysis.pdf [https://perma.cc/HK32-9N6A].

removal of inciting content that would allow harmful content to spread ("false negatives").[134]

C. Algorithmic-Based Targeting of Recommendations

> Our own information—from the everyday to the deeply personal—is being weaponized . . . . These scraps of data, each one harmless enough on its own, are carefully assembled, synthesized, traded and sold. Taken to the extreme this process creates an enduring digital profile and lets companies know you better than you may know yourself. Your profile is a bunch of algorithms that serve up increasingly extreme content, pounding our harmless preferences into harm.[135]

Hosting terrorists' content and providing communication tools to all users, or allowing all users to share revenues, is not the whole story. Intermediaries directly influence network dynamics from the top down.[136] To promote engagement, intermediaries make their website "sticky" causing users to become addicted to the engagement and keeping them on the website.[137] One way to do so is to amplify content that triggers strong emotional registers, including hate speech and extrem-

---

134. GILLESPIE, *supra* note 47, at 99, 107. Using algorithms to detect hate speech is likely to result in far more false positives than false negatives because algorithms cannot capture context—tone, speaker, and audience. Danielle Keats Citron & Neil M. Richards, *Four Principles for Digital Expression (You Won't Believe #3!)*, 95 WASH. U.L. REV. 1353, 1362 n.53 (2018).

135. Natasha Lomas, *Apple's Tim Cook makes blistering attack on the 'data industrial complex,'* TECHCRUNCH (Oct. 24, 2018, 5:24 AM), https://techcrunch.com/2018/10/24/apples-tim-cook-makes-blistering-attack-on-the-data-industrial-complex/ [https://perma.cc/RB4V-2M3P] (quoting Apple CEO Tim Cook) (internal quotation marks omitted).

136. *See* OLIVIER SYLVAIN, DISCRIMINATORY DESIGNS ON USER DATA 4 (2018), https://s3.amazonaws.com/kfai-documents/documents/28a74f6e98/Discriminatory-Designs-on-User-Data.pdf [https://perma.cc/82HZ-CVG7] ("Intermediaries today do much more than passively distribute user content or facilitate user interactions. Many of them elicit and then algorithmically sort and repurpose the user content and data they collect"); *see also* GILLESPIE, *supra* note 47, at 207 ("[P]latforms invoke and amplify particular forms of discourse and moderate away others . . . ."); O'Callaghan et al, *supra* note 36, at 460.

137. SHOSHANA ZUBOFF, THE AGE OF SURVEILLANCE CAPITALISM: THE FIGHT FOR A HUMAN FUTURE AT THE NEW FRONTIER OF POWER 466 (2019) (explaining that just as ordinary people can become compulsive gamblers at the hands of gaming industry, behavioral technology at the service of intermediaries draws ordinary people into an "unprecedented vortex of social information"); *see also* Karen Hao, *YouTube is experimenting with ways to make its algorithm even more addictive*, MIT TECH. REV. (Sept. 27, 2019), https://www.technologyreview.com/s/614432/youtube-algorithm-gets-more-addictive/ [https://perma.cc/E2N6-AHPU].

ism.[138] In fact, intermediaries are explicitly engineered to promote items that generate strong reactions because content that "evokes high-arousal emotion" is more likely to be shared, increase the engagement on the platform, and enhance the intermediary revenues.[139]

Intermediaries collect information on users, spy on them without consent,[140] and target them with personalized recommendations to connect with others, as more accurate recommendations enhance the attractiveness of the platform and increase users' engagement.[141] By "systemization of the personal," intermediaries influence, and even control, with whom users connect.[142] Intermediaries can also control what content users see online based on their past activities,[143] influence their feelings,[144] and cause them to consume more extreme

138. SIVA VAIDHYANATHAN, ANTISOCIAL MEDIA: HOW FACEBOOK DISCONNECTS US AND UNDERMINES DEMOCRACY 5–9 (2018) (describing how Facebook develops algorithms that favor highly charged content and depend on a self-serving advertising system that precisely targets ads using massive surveillance and personal dossiers).

139. Jonah Berger & Katherine L. Milkman, *What Makes Online Content Viral?*, 49 J. MARKETING RES. 192, 202 (2012); *see also* Rupert Neate, *Extremists made £250,000 from ads for UK brands on Google, say experts*, GUARDIAN (Mar. 17, 2017, 1:30 PM), https://www.theguardian.com/technology/2017/mar/17/extremists-ads-uk-brands-google-wagdi-ghoneim [https://perma.cc/JEF5-RZZW].

140. The recent documents leaked from Facebook demonstrate how companies spy on their users without consent. Sebastian Klovig Skelton & Bill Goodwin, *Lawmakers study leaked Facebook documents made public today*, COMPUTERWEEKLY.COM (Nov 6, 2019), https://www.computerweekly.com/news/252473540/Lawmakers-study-leaked-Facebook-documents-made-public-today [https://perma.cc/X5TD-Z27L].

141. *See* ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 84–85 (2018) (explaining that targeting combines information directly provided by a user, with data automatically generated from the use of the website, social media information, and data available from third parties to generate personalized information for each user); *see also* FRISCHMANN & SELINGER, *supra* note 64, at 150.

142. Ryan Calo, *Digital Market Manipulation*, 82 GEO. WASH. L. REV. 995, 1021 (2014).

143. Jack M. Balkin, *Free Speech is a Triangle*, 118 COLUM. L. REV. 2011, 2027 (2018) ("The creation of personalized feeds is inevitably content-based—social media sites have to decide what content is likely to be most interesting to its end-users."); *see also* Karen Levy & Solon Barocas, *Designing Against Discrimination in Online Markets*, 32 BERKELEY TECH. L.J. 1183, 1183, 1185–87 (2017).

144. Adam D.I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental evidence of massive-scale emotional contagion through social networks*, 111 PROC. NAT'L ACAD. SCI. 8788, 8788–90 (2014).

content.[145] Through this new form of "surveillance capitalism," intermediaries might predict and even engineer users' desires and behavior as a means to produce revenue.[146] Intermediaries also present advertisements to users. The advertisements are not placed randomly; instead, they are targeted to the viewer based on information harvested from the viewer's online behavior.[147] Targeting seems to improve in accuracy with AI development and usage of complex algorithms.[148]

Intermediaries use algorithmic targeting to improve user experiences and enhance engagement, but their practices are often problematic and result in techno-social engineering.[149] In contrast to hosting content and providing communication tools, personalizing content is an active and selective action of intermediaries that does not offer equal choice to all users. The intermediaries determine what recommendations, content, and advertisement will be available to whom. Thus, different people see different content and have different online experiences.[150]

Although it may appear that the system operates without human intervention, the intermediary structures it and the operation of the algorithm depends on the discretion of its programmers who can program it without neutrality or tinker with the results ex post.[151] Personalized algorithmic targeting is

---

145. YouTube's rabbit hole is the phenomenon of personalizing recommendations and playing recommended videos from a bottomless queue. *See* O'Callaghan et al., *supra* note 36, at 460; *see also* Hao, *supra* note 137.

146. FRISCHMANN & SELINGER, *supra* note 64, at 241 ("Our roles and desired can be engineered more than we appreciate."); S.C. Matz et al., *Psychological targeting as an effective approach to digital mass persuasion*, 114 PROC. NAT'L ACAD. SCI. 12714, 12714 (2017); Shoshana Zuboff, *Big other: surveillance capitalism and the prospects of an information civilization*, 30 J. INFO. TECH. 75, 83 (2015).

147. *See* Julia Angwin, Madeleine Varner & Ariana Tobin, *Facebook Enabled Advertisers to Reach 'Jew Haters*,*'* PROPUBLICA (Sept. 14, 2017, 4:00 PM), http://www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters [https://perma.cc/JV7L-PS2R].

148. On the influence of algorithms and AI on freedom of speech, see Jack M. Balkin, *Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation*, 51 U.C. DAVIS L. REV. 1149 (2018).

149. SYLVAIN, *supra* note 136, at 12; VAIDHYANATHAN, *supra* note 138, at 37, 55.

150. SYLVAIN, *supra* note 136, at 12.

151. *See* Michal Lavi, *Taking Out of Context*, 31 HARV. J.L. & TECH. 145, 154 (2017); Omer Tene & Jules Polonetsky, *Taming the Golem: Challenges of Ethical Algorithmic Decision-Making*, 19 N.C. J.L. & TECH. 125, 137–38 (2017). In a related context, it was revealed that Google's executives and engineers tinker with the search results without neutrality to favor specific business or to increase or decrease the visibility of specific types of content. Kirsten Grind et al., *How Google Interferes*

different from algorithms that depend on users' positive choices to search for content, because the algorithm targets specific users based on their characteristics.[152] Furthermore, even if a policy-neutral algorithm is used, the practice of targeting has self-reinforcing power. Algorithms are also never truly neutral.[153] This practice of algorithmic-based recommendations and targeting can influence users' future choices and the likelihood of changing their minds.[154] This influence may be positive or negative.[155] Beyond the general risk of infringement on users' autonomy and the risk of shackling them to their past interests and decisions,[156] intermediaries can present harmful content to specific users through an automated recommendation system.[157] Ysabel Gerrard demonstrates that by following terms related to the eating disorders bulimia and anorexia; she started getting more automatic recommendations for pro-eating-disorder content and suggestions for a list of users whose accounts she should follow.[158] Such recommendations can encourage bulimia and anorexia and result in self-harm. Likewise, YouTube has promoted "how to self-harm" tutorials for youngsters aged 13.[159] Some platforms ban content originated

---

With Its Search Algorithms and Changes Your Results, WALL STREET J. (Nov. 15, 2019, 8:15 AM), https://www.wsj.com/articles/how-google-interferes-with-its-search-algorithms-and-changes-your-results-11573823753 [https://perma.cc/3WMR-ERBT].

152. On "policy-neutral" versus "policy-directed" algorithms, *see* Tene & Polonetsky, *supra* note 151, at 137–38. Note that Facebook's advertising algorithm already uses categories of targeting that can promote hate speech. *See* Kerri A. Thompson, *Commercial Clicks: Advertising Algorithms as Commercial Speech*, 21 VAND. J. ENT. & TECH. L. 1019, 1020–21 (2019).

153. Pauline T. Kim, *Manipulating Opportunity*, 106 VA. L. REV. (forthcoming 2020) (manuscript at 4) ("*[E]ven if* an advertiser uses neutral targeting criteria and intends to reach a diverse audience, an ad targeting algorithm may distribute information about opportunities in a biased way.").

154. Michal S. Gal, *Algorithmic Challenges to Autonomous Choice*, 25 MICH. TECH. L. REV. 59, 61–63 (2018).

155. For example, information can be used to target content that discourages users from engaging in destructive behavior. *See, e.g.*, Hayley Tsukayama, *Facebook is using AI to try to prevent suicide*, WASH. POST (Nov. 27, 2017, 8:18 PM), https://www.washingtonpost.com/news/the-switch/wp/2017/11/27/facebook-is-using-ai-to-try-to-prevent-suicide/ [https://perma.cc/TPV3-3C5W].

156. ZUBOFF, *supra* note 137, at 329–45.

157. Gerrard, *supra* note 38, at 4498.

158. *Id.* at 4503–05.

159. *See* Sean Keach, *YouTube caught promoting deadly 'how to self harm' tutorials for youngsters aged 13*, SUN (Feb. 5, 2019, 12:28 PM), https://www.thesun.co.uk/tech/8356276/youtube-suicide-self-harm-videos [https://perma.cc/C35A-GDTM].

from conspiracy websites as a matter of policy, but at the same time promote conspiracy theories by algorithmic targeting.[160] Algorithmic targeting can incite and reinforce extremism because users that consume extremist content are more likely to get a recommendation to connect with affiliates of foreign terrorist organizations,[161] even if it is in direct opposition to the platform's own mechanism of control.[162]

Algorithmic personalization of inciting content can be damaging to society, because its narrowing of information reinforces users' prior dispositions and gets them to engage with more extreme connections and controversial ideas.[163] In fact, algorithms may push users to consume more inciting content and to connect with individuals with more radical beliefs, thus causing a self-feeding cycle in which terrorist content replicates itself.[164] Algorithmic incitement of an individual can have consequences on his social network by affecting his engagement with others and strengthening a feedback loop that enforces itself, amplifying ideological extremism, and pursuing viral spread.[165] Yet, in a response to economic imperatives, intermediaries are radically indifferent to the consequences.[166]

Can words kill? Prosecutors seem to answer this question with a resounding "yes," as demonstrated by a recent case in which a woman was found guilty of involuntary manslaughter after encouraging her boyfriend to commit suicide via text messages and a phone call.[167] "Words can kill" is not an abstract notion. Inciting words have tangible and long-term ef-

---

160. *See* Manheim & Kaplan, *supra* note 38, at 147.

161. *See* Schwartz, *supra* note 106, at 1208–09 (explaining that Facebook's data-usage-policy algorithm can connect terrorists with sympathizers and other terrorists).

162. Gerrard, *supra* note 38, at 4505–06.

163. ELI PARISER: THE FILTER BUBBLE: WHAT THE INTERNET IS HIDING FROM YOU 35–48 (2011).

164. Schwartz, *supra* note 106, at 1209.

165. JULIE E. COHEN, BETWEEN TRUTH AND POWER: THE LEGAL CONSTRUCTIONS OF INFORMATIONAL CAPITALISM 76, 85 (2019).

166. ZUBOFF, *supra* note 137, at 505, 509.

167. *See* Commonwealth v. Carter, 115 N.E.3d 559, 561–65, 574 (Mass. 2019), *cert. denied sub nom.* Carter v. Massachusetts, 140 S. Ct. 910 (2020); Daniel Etcovitch, *Commonwealth v. Michelle Carter: Involuntary Manslaughter Conviction for Encouraging Suicide Over Text and Phone*, JOLT DIGEST (June 25, 2017), https://jolt.law.harvard.edu/digest/commonwealth-v-michelle-carter-involuntary-manslaughter-conviction-for-encouraging-suicide-over-text-and-phone [https://perma.cc/ZKB2-FX3V].

fects. Recent research demonstrates that words may lead to criminal behavior.[168] Terrorists' propaganda can also be linked to actual terrorist incidents.[169] But do platforms kill? Should intermediaries be held responsible for terrorists' attacks? Should they be liable for terrorist content on their platforms? Should they bear responsibility for incitement caused by an algorithm?

Because platforms operate differently, their liability must correspond with the action they have taken. It would be inappropriate to evaluate their liability according to a uniform standard; instead, their liability for aiding terrorism should be proportional to their activity, whether they host content or use algorithms. For this reason, it is vital to map and understand the roles intermediaries play online to make rational legal policy.

III. THE NEW SCHOOL OF REGULATION: INTERMEDIARIES' LIABILITY TO TERROR CONTENT

In traditional, or what Professor Jack Balkin calls old-school speech regulation, states imposed imprisonment or fines to regulate or control speech.[170] This is a "*dualist* or *dyadic* system of speech regulation."[171] In this model, there are essentially two players: the state and the speaker.[172] In contrast, the twenty-first century has created a pluralist model, what Professor Balkin calls new-school regulation, with many different players.[173] This model can be condensed into a triangle of actors: the state, the infrastructure, and the speaker.[174] Whereas old-school regulation is directed at speakers, new-school speech regulation is

168. *See* Raphael Cohen-Almagor, *Taking North American White Supremacist Groups Seriously: The Scope and Challenge of Hate Speech on the Internet*, 7 INT'L J. CRIME JUST. & SOC. DEMOCRACY, June 2018, at 38, 38–39; *see also* Simon Cottee, *Can Facebook Really Drive Violence?*, ATLANTIC (Sept. 9, 2018), https://www.theatlantic.com/international/archive/2018/09/facebook-violence-germany/569608/ [https://perma.cc/3KMK-PD97].

169. *See, e.g.*, Foo Yun Chee, *EU parliament votes to fine internet firms for not removing extremist content quickly*, REUTERS (Apr. 17, 2019, 3:03 PM), https://reut.rs/2WIJkPy [https://perma.cc/5ZQF-2H7G].

170. Balkin, *supra* note 143, at 2015.

171. *Id.* at 2013.

172. *Id.*

173. *Id.* at 2014.

174. *Id.* at 2014–15.

also directed at the infrastructure,[175] such as internet platforms that are private actors and currently not bound by the First Amendment.[176] States (or supernational entities like the European Union) attempt to regulate, threaten, coerce or co-opt elements of key players that shape the internet in order to get their infrastructure to surveil, police, and control speakers.[177] As digital infrastructure companies become increasingly more powerful in governing their spaces and collecting and analyzing content from their end users, states demand more from these companies through new-school speech regulation,[178] in an attempt to incentivize cooperation from the private sector.[179]

The following Part reviews the legal response to terrorists' speech. It refers to the normative considerations taken into account in the new-school form of regulation, which imposes liability and obligations on intermediaries for terrorists' speech.

A. Legal Response to Terrorist's Content on Social Media

1. Terrorists' Content Regulation in the Shadow of the Law

A consensus exists regarding the dangers posed by terrorists' use of the internet.[180] Governments around the world have recognized this threat and have started forcing intermediaries to remove unlawful material from their platforms.[181] Online intermediaries such as Facebook, Google, and Twitter have been

175. *Id.* at 2015.

176. *See, e.g.*, Prager University v. Google LLC, No. 18-15712, 2018 WL 913661, at *1 (9th Cir. Feb. 26, 2020) ("Despite YouTube's ubiquity and its role as a public-facing forum, it remains a private forum, not a public forum subject to judicial scrutiny under the First Amendment.").

177. *See* Jack M. Balkin, *Old-School/New-School Speech Regulation*, 127 HARV. L. REV. 2296, 2297–99 (2014).

178. *See id*.

179. O'Leary, *supra* note 68, at 559–60.

180. Tsesis, *supra* note 69, at 675–84.

181. DAPHNE KELLER, WHO DO YOU SUE? STATE AND PLATFORM HYBRID POWER OVER ONLINE SPEECH (Hoover Inst., Aegis Series Paper No. 1902, 2019), https://www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf [https://perma.cc/RW6R-BKVG]; Brian Chang, *From Internet Referral Units to International Agreements: Censorship of the Internet by the UK and EU*, COLUM. HUM. RTS. L. REV., Winter 2018, at 114, 116–20.

threatened with litigation in Australia,[182] Israel,[183] Germany,[184] France,[185] Spain,[186] the United Kingdom,[187] the EU,[188] and many other jurisdictions. The European Commission adopted measures to effectively tackle unlawful online content beyond takedown notices, as it took a more proactive approach towards terrorist content.[189] Recently, the European Parliament also voted to fine firms like Facebook, Google, and Twitter up to four percent of their turnover if they persistently fail to re-

182. *See* Evelyn Douek, *Australia's New Social Media Law Is a Mess*, LAWFARE (Apr. 10, 2019, 8:28 AM), https://www.lawfareblog.com/australias-new-social-media-law-mess [https://perma.cc/P8BK-TKPQ].

183. In Israel a new social media legislative bill would enable courts to order social networks to remove posts that are not under Israel's jurisdiction. Klein & Flinn, *supra* note 17, at 82. As a result, social networks might block more accounts related to terror organizations. *See Hezbollah says some of its Facebook and Twitter pages shuttered*, TIMES ISRAEL (June 23, 2018, 11:10 AM), https://www.timesofisrael.com/facebook-twitter-pages-of-hezbollah-shuttered/ [https://perma.cc/T8A7-N7DL]. But this bill ultimately did not pass. Gil Hoffman, *Netanyahu halts Facebook bill at last-minute*, JERUSALEM POST (July 18, 2018 12:39 PM), https://www.jpost.com/Israel-News/Netanyahu-halts-Facebook-bill-at-last-minute-562802 [https://perma.cc/8LMS-68HY].

184. *See* Chang, *supra* note 181, at 116–18; Reuters in Frankfurt, *German Facebook boss to be investigated for 'ignoring racist posts,'* GUARDIAN (Nov. 10, 2015, 2:07 PM), https://www.theguardian.com/technology/2015/nov/10/german-facebook-boss-investigated-hamburg-prosecutors-hate-speech [https://perma.cc/556X-EPLM].

185. Lizzie Plaugic, *France wants to make Google and Facebook accountable for hate speech*, VERGE (Jan. 27, 2015, 12:38 PM), https://www.theverge.com/2015/1/27/7921463/google-facebook-accountable-for-hate-speech-france [https://perma.cc/K95P-LLAW].

186. *See* Roter, *supra* note 101, at 1400–01.

187. *See* Terrorism Act 2006, c.11. Under this law, "platforms have only two days to comply with a takedown request; otherwise, they are deemed to have 'endorsed' the terrorist content." *See* GILLESPIE, *supra* note 47, at 37. In addition, the Counter-Terrorism and Border Security Act of 2019 updates terrorism offences for the digital age and grants the authorities more power to tackle the threat posed to the United Kingdom by terrorism. Counter-Terrorism and Border Security Act 2019, c.3.

188. *See* Chang, *supra* note 181, at 117–18; Liat Clark, *Facebook and Twitter must tackle hate speech or face new laws*, WIRED UK (Dec. 5, 2016), https://www.wired.co.uk/article/us-tech-giants-must-tackle-hate-speech-or-face-legal-action [https://perma.cc/4JYG-RPFA]; *see also* GILLESPIE, *supra* note 47, at 121 ("[I]n 2016 all of the major platforms promised European lawmakers to ensure review of possible terrorist or extremist content within a one-day window.").

189. *See* Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on certain legal aspects of information society services, in particular electronic commerce in the Internal Market ('Directive on electronic commerce'), arts. 14, 15, 16, 2000 O.J. (L 178) 13, 14; Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online, 2018 O.J. (L 63) 52.

move extremist content within one hour of being asked to do so by authorities.[190]

"On May 31, 2016, Facebook, Microsoft, Twitter, and YouTube entered into an agreement with the European Commission to remove 'hateful' speech within twenty-four hours if appropriate under terms of service."[191] Technology companies also contemplated trying to establish a database for detecting banned violent terrorist images, audio, and video files.[192] The database was supposed to include unique digital fingerprints of banned content so that files could be flagged and removed instantly.[193] At first, technology companies rejected the idea; however, six months later, they announced plans for an industry database "to help prevent the spread of violent terrorist imagery."[194] The tech companies issued guidelines that limited the use of the database for the most extreme terrorists' images that violate the content policies of all companies.[195] Furthermore, according to the guidelines, the removal of hashed material would not be automatic but rather subjected to a review by the tech company according to its own specific policies.[196]

Recently, the European Parliament issued a proposal for a regulation for preventing the dissemination of terrorist content online.[197] The regulation proposed obligations to prevent unlawful content from reappearing after its removal.[198] Similarly, in a related context, the European Court of Justice held that EU law does not preclude intermediaries such as Facebook from being ordered to remove identical and, in certain circumstances, equivalent comments previously declared unlawful.[199] Thus, it seems that legal obligations for more proactive moderation are becoming more widespread in the EU.

---

190. *See* Chee, *supra* note 169.

191. Citron, *supra* note 132, at 1038.

192. *Id.* at 1043.

193. *Id.* at 1043–44.

194. *Id.* at 1044–45.

195. *Id.* at 1045.

196. *Id.* at 1045.

197. *See* EUR. PARL. DOC. (COM 0640) (2018).

198. *See id.* at 26.

199. *See* Case C-18/18, Eva Glawischnig-Piesczek v Facebook Ireland Limited 2019 EUR-Lex CELEX LEXIS 458 (Oct. 3 2019). For criticism on the Advocate General's opinion, see KELLER, *supra* note 133, at 15–28.

The Counter Terrorism Internet Referral Units (CTIRU), which was created in the United Kingdom but was adopted in a number of countries, followed a policy of ad hoc, ex post removal policy.[200] Given the success of the CTIRU's efforts, Europol established its Internet Referral Unit, describing them as a "partnerships with the private sector."[201] Ninety-one percent of the content reported by the unit has been removed.[202]

Clearly, governments' threat of legislation incentivizes intermediaries to invest resources in reducing terrorists' content. Although this dynamic has the potential to reduce terrorist content, it does not address algorithmic targeting nor lead to compensation of terror victims.

2. The U.S. Approach

In the United States, the First Amendment grants extensive protection to freedom of speech and restricts government from constraining speech.[203] Thus, there is a presumption against content-based speech restrictions.[204] Despite the threat of terrorists using social media, different regulatory initiatives to impose obligations on intermediaries regarding terrorists' content have not been enacted by Congress or adopted by the industry.[205] The current law does not impose specific positive obligations on online intermediaries for preventing terrorist's content on their platforms. Nevertheless, victims of terror attacks have filed claims under the civil enforcement provisions of the fed-

200. *See* Chang, *supra* note 181, at 120–22.

201. EUROPOL, EU INTERNET REFERRAL UNIT: YEAR ONE REPORT HIGHLIGHTS 3 (2016), https://www.europol.europa.eu/sites/default/files/documents/eu_iru_1_year_report_highlights.pdf [https://perma.cc/P33K-9HRN].

202. *Id.* at 5.

203. U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech, or of the press.").

204. Ashcroft v. ACLU, 542 U.S. 656, 660 (2004) (citing R.A.V. v. City of St. Paul, 505 U.S. 377, 382 (1992)).

205. *See, e.g.*, Intelligence Authorization Act for Fiscal Year 2016, H.R. 2596, 114th Cong. (2015); Requiring Reporting of Online Terrorist Activity Act, S. 2372, 114th Cong. (2015); GILLESPIE, *supra* note 47, at 39 (referring to the "Obama administration urg[ing] tech companies to develop new strategies for identifying extremist content, either to remove it, or report it to the national security authorities"); Citron, *supra* note 132, at 1036 (referring to Senator Lieberman's demand from media giants to remove terrorists' content).

eral antiterrorism laws, basing their theory of liability on material support doctrines.[206]

a. Material Support Doctrines

Section 2339A of the United States Code prohibits one from providing "material support or resources . . . knowing or intending that they are to be used in preparation for, or in carrying out" a violation of certain offenses, including terrorism.[207] Unlike § 2339A, § 2339B does not include a knowing or intentional mens rea element, or specific intent, but rather prohibits the willful provision of anything of value to a group designated as a Foreign Terrorist Organization (FTO).[208] Thus, if a provider knows that an organization has been officially designated as "terrorist," or if it knows that an organization engages in terrorism, it may be found guilty. The lack of a specific intent requirement under § 2339B has been a persistent source of criticism.[209]

In *Holder v. Humanitarian Law Project* (*HLP*),[210] the Supreme Court upheld the constitutionality of 2339B, and determined that the federal government had the authority to prohibit groups from working with terrorist organizations even when their violent operations were interlinked with more benign functions, such as charity work.[211] Because of the grave danger posed by terrorist organizations, the Supreme Court interpreted coordination in broad terms, determining that working in coordination with or at the command of FTOs serves to legitimize and further their terrorist means, and therefore these actions

206. *See, e.g.*, *Crosby* Complaint, *supra* note 7, at 49–50.

207. 18 U.S.C. § 2339A (2018); *see also* Schwartz, *supra* note 106, at 1186.

208. 18 U.S.C. § 2339B (2018); *see also* Schwartz, *supra* note 106, at 1186. An FTO is an organization that the Secretary of State has designated to be foreign terrorists. *See id.* The list of FTOs maintained by the State Department encompasses sixty-one such groups. Bureau of Counterterrorism, U.S. Dep't of State, *Foreign Terrorist Organizations* (last visited Nov. 12, 2019), http://www.state.gov/j/ct/rls/other/des/123085.htm [https://perma.cc/ZG7U-W93P].

209. *See, e.g.*, David Cole, *Out of the Shadows: Preventive Detention, Suspected Terrorists, and War*, 97 CALIF. L. REV. 693, 724–25 (2009); Rachel E. VanLandingham, *Jailing the Twitter Bird: Social Media, Material Support to Terrorism and Muzzling the Modern Press*, 39 CARDOZO L. REV. 1, 48 (2017).

210. 561 U.S. 1 (2010).

211. *Id.* at 7–8. In *HLP*, the plaintiffs sought to provide training in international law, political involvement, and negotiation strategies to Partiya Karkeran Kurdistan and the Liberation Tigers of Tamil Eelam. *Id.* at 9, 14–15. Both groups are on the State Department's designated terrorist organization list. *Id.* at 9. The material support, however, was not directly linked to illegal terrorists' actions. *Id.* at 16.

are considered material support.[212] D.C. District Court Judge Collyer recently ordered Iran and Syria to pay $4.1 million in damages to a family of terror victims, because these countries materially supported and gave resources to Hamas in Israel, which contributed to the hostage taking and murder of a 16-year-old victim.[213]

Sections 2339A and 2339B do not create a private civil cause of action, but § 2333 "allows private parties who are nationals of the United States to sue in federal district court and receive treble damages and attorney's fees if they were injured in their 'person, property, or business by reason of international terrorism.'"[214] The scienter requirement "may be satisfied when an entity recognizes it is supporting a terrorist organization; it needs not be aware that its aid is going to advance a specific terrorist conspiracy."[215]

In September 2016, Congress enacted the Justice Against Sponsors of Terrorism Act (JASTA),[216] which expanded anti-terrorism law by adding 18 U.S.C. § 2333(d).[217] JASTA provides that U.S. nationals may assert liability against a person who aids and abets or conspires with a person who commits an act of international terrorism.[218]

"Since its enactment, section 2333 has primarily targeted financial institutions, banks, and charitable organizations that provide material support in the form of fundraising to FTOs."[219] Following the terrorist attacks in the last three years,

---

212. *Id.* at 30–31; *see also* United States v. Mehanna, 735 F.3d 32, 46, 49–50 (1st Cir. 2013) (holding that sufficient evidence of "coordination" existed where Mehanna had merely attempted to travel to an al-Qaeda training camp). For criticism of the *HLP* decision, see David Cole, *The First Amendment's Borders: The Place of* Holder v. Humanitarian Law Project *in First Amendment Doctrine*, 6 HARV. L. & POL'Y REV. 147 (2012).

213. Elisha Ben Kimon, *US court blames murdered teen's family for living in territories*, YNET NEWS (July 3, 2018 9:57 AM), https://www.ynetnews.com/articles/0,7340,L-5302974,00.html [https://perma.cc/LEJ4-ABE2].

214. Klein & Flinn, *supra* note 17, at 85 (quoting 18 U.S.C. § 2339B (2018)).

215. Tsesis, *supra* note 17, at 620.

216. Pub. L. No. 114-222, 130 Stat. 852 (2016) (codified in scattered sections of 18 and 28 U.S.C.).

217. *Id.* § 4, 130 Stat. at 854.

218. *See* Jennifer Steinhauer, Mark Mazzetti & Julie Hirschfeld Davis, *Congress Votes to Override Obama Veto on 9/11 Victims Bill*, N.Y. TIMES (Sept. 28, 2016), https://nyti.ms/2dkxCaB [https://perma.cc/J7H7-LV2E].

219. Schwartz, *supra* note 106, at 1088.

however, § 2333 has become the basis for civil cases against social media companies.[220] Family members of terror victims argue that social media companies knowingly cooperate with designated foreign terrorists in posting, displaying, or hosting propaganda, which have resulted in the deaths of American nationals.[221] Section 230 and other legal requirement, however, pose challenges for plaintiffs.

b. Section 230 of the Communication Decency Act

Section 230(c)(1) of the Communications Decency Act (CDA)[222] reflects the strong U.S. bias favoring free speech over other values.[223] Under the subsection heading "Protection for 'Good Samaritan' blocking and screening of offensive material," it directs, "[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."[224] In passing § 230, Congress sought to promote self-regulation, free speech, and foster the rise of vibrant internet enterprises.[225] Thus, a defendant that provides a forum for communicating materials is not likely to be responsible as a content provider.[226]

---

220. *Id.* at 1089.

221. *Id.* at 1189; *see* Freilich, *supra* note 10, at 677.

222. 47 U.S.C. § 230 (2018).

223. JEFF KOSSEFF, THE TWENTY-SIX WORDS THAT CREATED THE INTERNET 246 (2019); Eric Goldman, *Why Section 230 Is Better than the First Amendment*, 95 NOTRE DAME L. REV. REFLECTION 33 (2019).

224. 47 U.S.C. § 230(c)(1).

225. *See* 47 U.S.C. § 230(b)(1)–(2); Anupam Chander, *How Law Made Silicon Valley*, 63 EMORY L. J. 639, 651–52 (2014).

226. Cecilia Ziniti, *The Optimal Liability System for Online Service Providers: How* Zeran v. America Online *Got it Right and Web 2.0 Proves It*, 23 BERKELEY TECH. L.J. 583, 585 (2008) ("Almost uniformly, courts have interpreted § 230's safe harbor broadly."). There are, however, some exceptions to the immunity. It is limited to civil claims and does not apply to cases that are based on federal criminal laws. 47 U.S.C. § 230(c)(2). In addition, there have been legislative efforts to narrow this defense for specific types of speech. *See* Allow States and Victims to Fight Online Sex Trafficking Act of 2017, Pub. L. No. 115-164, 132 Stat. 1253; *see also* Mary Graw Leary, *The Indecency and Injustice of Section 230 of the Communications Decency Act*, 41 HARV. J.L. & PUB. POL'Y 553, 556–57 (2018); Eric Goldman, *'Worst of Both Worlds' FOSTA Signed Into Law, Completing Section 230's Evisceration*, TECH. & MARKETING L. BLOG (Apr. 11, 2018), https://blog.ericgoldman.org/archives/2018/04/worst-of-both-worlds-fosta-signed-into-law-completing-section-230s-evisceration.htm [https://perma.cc/2LK9-RSWX].

Courts have interpreted § 230 broadly and repeatedly shielded web enterprises from lawsuits in a plethora of cases.[227]

Section 230 applies to secondary liability. However, if the intermediary is "responsible" in whole or in part for the "creation or development" of content, courts may find the intermediary liable as an information content provider.[228] Section 230 does not define "creation" or "development." Thus, the line between the service itself and the creation of information is blurry and the scope of liability is ambiguous.[229] In the beginning, courts applied the immunity in nearly all cases.[230]

A decade ago, an important case, *Fair Housing Council of San Fernando Valley v. Roommates.com, LLC*,[231] led to confusion regarding intermediaries' liability. That case dealt with a website that allowed users to find roommates, Roommates.com.[232] The website required users to fill in a personal profile and answer several questions, including questions about the users' genders and sexual orientations, and to express their preferences on these issues with respect to roommates.[233] The answers were chosen from check box and drop down menus.[234] An internal search engine allowed users to search roommates while filtering unfit matches according to these criteria.[235] The website also

---

227. *See* Chander, *supra* note 225, at 653; Lavi, *supra* note 48, at 867–70; *see also* Herrick v. Grindr LLC, 765 F. App'x 586, 589 (2d Cir. 2019); Zeran v. Am. Online, Inc., 129 F.3d 327, 330 (4th Cir. 1997) ("By its plain language, § 230 creates a federal immunity to any cause of action that would make service providers liable for information originating with a third-party user of the service."); Caraccioli v. Facebook, Inc., 167 F. Supp. 3d 1056, 1066 (N.D. Cal. 2016) (holding that the immunity applies even when the intermediary knew of the defamatory content, reviewed it, and decided not to remove it), *aff'd*, 700 F. App'x 588 (9th Cir. 2017).

228. Anupam Chander & Uyén P. Lé, Free *Speech*, 100 IOWA L. REV. 501, 514 (2015); Zak Franklin, *Justice for Revenge Porn Victims: Legal Theories to Overcome Claims of Civil Immunity by Operators of Revenge Porn Websites*, 102 CALIF. L. REV. 1303, 1316 (2014).

229. *See* Ken S. Myers, *Wikimmunity: Fitting the Communications Decency Act to Wikipedia*, 20 HARV. J.L. & TECH. 163, 201 (2006).

230. *See* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 461–62 (2010); Freilich, *supra* note 10, at 683.

231. 489 F.3d 921 (9th Cir. 2007), *aff'd in part, rev'd in part, vacated in part en banc*, 521 F.3d 1157 (9th Cir. 2008).

232. *Id.* at 924.

233. *Id.* at 924, 926.

234. *Id.* at 926.

235. *Id.* at 928–29.

included an open section for users' comments.[236] The intermediary sent periodical emails to users, which included only potential matches.[237] The Fair Housing Council argued that the questions in the drop down menus violated the federal Fair Housing Act[238] and led to discrimination.[239]

The first court to consider the issue dismissed the case because of § 230 immunity.[240] On appeal, the Ninth Circuit declined to grant Roommates.com immunity.[241] The en banc rehearing majority opinion reached the same conclusion.[242] Writing for the majority, Chief Judge Kozinsky stressed that although the Communications Decency Act created immunity, it "was not meant to create a lawless no-man's-land on the Internet."[243] The en banc court held that the intermediary provided a limited set of pre-populated discriminatory answers and required users to choose.[244] The court determined that Roommates.com was an information content provider with respect to the illegal housing discriminatory questions on the site.[245] An information content provider is "more than a passive transmitter of information."[246] The court also declined to grant immunity for the internal search engine and the email mechanism because both did not use neutral tools but instead channeled the distribution of discriminatory content.[247] However, the court held immunity applied to materials posted in the open comment section.[248]

---

236. *Id.* at 924.

237. *Id.* at 928–29.

238. 42 U.S.C. § 3604(c) (2018).

239. *Roomates.com*, 489 F.3d at 924.

240. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, No. CV 03-09386PA(RZX), 2004 WL 3799488, at *6 (C.D. Cal. Sept. 30, 2004).

241. *Roommates.com*, 489 F.3d at 926.

242. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC, 521 F.3d 1157, 1175 (9th Cir. 2008) (en banc).

243. *Id.* at 1164.

244. *Id.* at 1164–67.

245. *Id.* at 1164.

246. *Id.* at 1166.

247. *Id.* at 1167.

248. The court reasoned that:

> A website operator can be both a service provider and a content provider: If it passively displays content that is created entirely by third parties, then it is only a service provider with respect to that content. But as to content that it creates itself, or is "responsible, in whole or in part" for creating or developing, the website is also a content provider. Thus, a

After *Roommates.com*, courts expressed doubts regarding the scope of immunity, resulting in many contradictory judicial decisions. For example, in *Dyroff v. Ultimate Software Group, Inc.*,[249] the court held that immunity applied even when the intermediary used data mining and machine learning algorithms that allowed the provider to analyze data on users and to channel users to participate in particular groups and consume particular types of content.[250]

On appeal, the Ninth Circuit affirmed the lower court, concluding that by recommending user groups and sending email notifications, Ultimate Software, through its Experience Project website, was acting as a publisher of others' content.[251] These functions—recommendations and notifications—are tools meant to facilitate the communication and content of others and are not content in and of themselves.[252] The court concluded that Ultimate Software's functions on Experience Project most resemble the "Additional Comments" features in *Roommates.com*.[253] The recommendation and notification functions helped facilitate this user-to-user communication, but it did not materially con-

---

website may be immune from liability for some of the content it displays to the public but be subject to liability for other content.

Id. at 1162–63.

249. No. 17-cv-05359-LB, 2017 WL 5665670 (N.D. Cal. Nov. 26, 2017), *aff'd*, 934 F.3d 1093 (2019).

250. *Id.* at *1–2. Data mining and machine learning allowed the intermediary to personalize recommendations to users regarding content and discussion groups that might be of interest for the user. *Id.* In some cases, the recommendations channeled users to unlawful content. *Id.* at *9. In one instance, the recommendations allegedly steered a user to a discussion group dedicated to the sale of narcotics. *Id.* at *4. The communication on the website allegedly allowed the user to buy heroin and he died because he consumed it. *Id.* at *1. The court dismissed the case ruling that recommendations to users are ordinary, neutral functions of social network websites. *Id.* at *9. The intermediary used neutral tools that merely provided a framework that could be utilized for proper or improper purposes. *Id.* at *10. As such, it did not "create or develop" the information even in part. *Id.* at *1. Therefore, the immunity applied. *Id.* The situation in *Dyroff* is similar to the email service in *Roommates.com*. The court could reach a different conclusion because the platform gained new information from users' content and behavior to create a site architecture that affects behavior.

251. *Dyroff*, 934 F.3d at 1094, 1096.

252. *Id.* at 1096.

253. *Id.* at 1099.

tribute to the alleged unlawful content.[254] Dyroff appealed to the Supreme Court.[255]

A similar narrow interpretation regarding algorithmic contribution to unlawful content was also adopted in the related context of intellectual property.[256] In contrast, in *Daniel v. Armslist, LLC*,[257] the Wisconsin Court of Appeals interpreted *Roommates.com* broadly and did not grant immunity for website features that facilitated the purchase of illegal firearms that were used in a fatal shooting.[258] The court did not grant immunity even though only some of the transactions ended up being illegal on the buyer's side.[259] On appeal, the Supreme Court of Wisconsin reversed the decision of the appellate court, reasoning that the defendant provided neutral tools that could be used for lawful purpose, and third parties used them to create unlawful content.[260] The court also explained that § 230 does not contain a good faith requirement.[261] Liability is derived from the intermediary's function as publisher or speaker.[262] Thus, according to the Wisconsin Supreme Court, the immunity applies even if the intermediary has knowledge of unlawful content on its platform, and even if it designs the website to facilitate unlawful activity by not including phone or email verification.[263]

---

254. *Id.*

255. Petition for a Writ of Certiorari, Dyroff v Ultimate Software, No. 19-849 (U.S. Jan. 02, 2020).

256. *See* Ventura Content, Ltd. v. Motherless, Inc., 885 F.3d 597, 605–08 (9th Cir. 2018) (concluding that neither tagging, group, nor algorithmic suggestions for "most popular" material change the user-submitted status of the material).

257. 913 N.W.2d 211 (Wis. Ct. App. 2018), *rev'd*, 926 N.W.2d 710 (Wis. 2019).

258. *Id.* at 214, 224.

259. *See id.* at 215. The opinion does not detail the exact circumstances when its statutory reading would support a section 230 defense. *See* Eric Goldman, *Wisconsin Appeals Court Blows Open Big Holes in Section 230—Daniel v. Armslist*, TECH. & MARKETING L. BLOG (Apr. 25, 2018), https://blog.ericgoldman.org/archives/2018/04/wisconsin-appeals-court-blows-open-big-holes-in-section-230-daniel-v-armslist.htm [https://perma.cc/E9VM-U6U4].

260. *Armslist*, 926 N.W.2d at 721–22, 727, *cert. denied*, 140 S. Ct. 562 (2019).

261. *Id.* at 722.

262. *Id.* at 723–24.

263. *See id.* at 726 ("That Armslist may have known that its site could facilitate illegal gun sales does not change the result. Because § 230(c)(1) contains no good faith requirement, courts do not allow allegations of intent or knowledge to defeat a motion to dismiss."). The plaintiff filed a petition to the United States Supreme Court on this case, but the petition was denied. *Armslist*, 140 S. Ct. 562.

In general, courts choose to err on the side of immunity; however, the exact standards for excluding intermediaries from immunity remain unclear.

c. Challenges to Civil Lawsuits under Sections 2333 and 230 and Proximate Cause

Sections 2339A and 2339B are criminal provisions. However, § 2333 allows a plaintiff to file civil suits. Thus, courts have concluded that § 230 applies to civil claims based on federal crimes.[264] Defendants in civil litigation argue that they did not publish the content and therefore, they are not responsible for something they did not produce.[265] Most courts have granted media companies' motions to dismiss based on § 230, rejecting civil suits even when plaintiffs based their claim on direct liability and involvement of the intermediaries in the creation of information and targeting of offending messages.[266]

Another challenge to suits under § 2333 is the requirement for causal connection between the conduct and the injury under the material support statues.[267] This is a well-established principle of common law in torts. The requirement for causal connection is reflected in the words "by reason of" in § 2333(a).[268] This requirement is not but-for causation, but different circuits have different standards for establishing proximate cause.[269] The Seventh Circuit maintains a relatively low standard: there must be a substantial probability that a service is a contributing cause of an attack.[270] In contrast, the Second

---

264. *See, e.g.*, Cohen v. Facebook, Inc., 252 F. Supp. 3d 140, 157–58 (E.D.N.Y. 2017).

265. *See* Schwartz, *supra* note 106, at 1192.

266. *See e.g.*, Crosby v. Twitter, 303 F. Supp. 3d 564, 574–75 (E.D. Mich. 2018).

267. Schwartz, *supra* note 106, at 1200–02.

268. 18 U.S.C. § 2333(a) (2018) ("Any national of the United States injured in his or her person, property, or business *by reason of* an act of international terrorism . . . ." (emphasis added)); *see also* Fields v. Twitter, 881 F.3d 739, 744–45 (9th Cir. 2018) ("'If, in creating civil liability through § 2333, Congress had intended to allow recovery upon a showing lower than proximate cause, we think it either would have so stated expressly or would at least have chosen language that had not commonly been interpreted to require proximate cause for the prior 100 years' . . . . In light of these assumptions, we understand that the phrase 'by reason of' connotes some degree of directness." (alteration adopted) (quoting Rothstein v. UBS AG, 708 F.3d 82, 95 (2013))).

269. Brown, *supra* note 122, at 26–27.

270. *Id.* at 27; *see also* Boim v. Holy Land Found. for Relief & Dev., 549 F.3d 685, 697 (7th Cir. 2008) (en banc); Schwartz, *supra* note 106, at 1201.

Circuit requires that a terrorist attack be a foreseeable consequence of the specific act of support,[271] and the Ninth Circuit requires direct relation between the platform and the plaintiffs' injury.[272]

As the following subsections demonstrate, courts have dismissed civil cases based on material support statutes against media giants. *Cohen v. Facebook, Inc.*[273] involved two sets of claims filed by the "Force Plaintiffs" and the "Cohen Plaintiffs" against Facebook that focus on the presence of the Palestinian terrorist group Hamas in social media.[274] The Cohen Plaintiffs, 20,000 Israeli citizens, filed a negligence suit, asserting that Palestinian terrorists used Facebook to incite, enlist, and organize would be killers to slaughter Jews.[275] Facebook allegedly knowingly allowed terror organizations to operate Facebook accounts using their own names, and Facebook's approach towards removal was inconsistent.[276] The plaintiffs further alleged that Facebook's algorithms connected users with other users, groups, and content that might interest them.[277] This recommendation system played a vital role in spreading terrorist content to those who were most susceptible to the message and likely to act upon the incitement.[278] Because of wild incitement on social media, the plaintiffs argued that future attacks threaten them.[279] The case was dismissed for lack of standing because the individual plaintiffs asserted only a threat or fear of possible future harm, which was not "actual or imminent."[280]

The Force Plaintiffs, family members of victims of terrorists' attacks in Israel, filed a suit against Facebook. The plaintiffs based their claim on the material support doctrine, asserting that Facebook was liable for its own content that was not gen-

---

271. *See Rothstein*, 708 F.3d at 91.

272. *Fields*, 881 F.3d at 746, 750.

273. 252 F. Supp. 3d 140 (E.D.N.Y. 2017), *aff'd sub nom.* Force v. Facebook, Inc., 934 F.3d 53, 76 (2d Cir. 2019).

274. *Id.* at 146, 157–58.

275. *Id.* at 146.

276. *Id.* at 146–47.

277. *Id.*

278. *Id.*

279. *Id.* at 146.

280. *Id.* at 150 ("A plaintiff alleging only an 'objectively reasonable possibility' that it will sustain the cited harm at some future time does not satisfy this requirement." (quoting Clapper v. Amnesty Int'l USA, 568 U.S. 398, 409–10 (2013))).

erated by another, because Facebook provided "network[ing]" and "broker[ed]" links among terrorists.[281]

The court dismissed the case under § 230.[282] It ignored the legal problem of inciting-content recommendations and concluded that there was no difference between making the system available to terrorists and providing terrorists with valuable services.[283] The services are part and parcel of access to a Facebook account and so imposing liability on that basis would turn on "Facebook's choices as to who may use its platform."[284] The court further reasoned that the features criticized by the plaintiffs operate solely in conjunction with content posted by Facebook users.[285] Thus, the court rejected the case and denied the plaintiffs' request to reconsider the ruling.[286]

On appeal to the Second Circuit, the Force Plaintiffs argued that the district court "improperly dismissed their claims because Section 230(c)(1) does not provide immunity to Facebook under the circumstances of their allegations."[287] They argued that providing a forum for communication for terrorists, facilitating personalized "newsfeed" pages for each user, and providing "friends suggestions" by using algorithms extend beyond a function of an information content provider.[288] They argued that, in fact, Facebook is acting as a publisher of the information, and even provides the content by itself, by targeting it with algorithms and contributing to terrorists' content.[289]

The Second Circuit concluded that the district court properly applied § 230(c)(1) to plaintiffs' federal claims.[290] The court determined that § 230 should be read broadly.[291] Giving Hamas a forum for communication falls under § 230(c)(1), because

---

281. Force Complaint at 2, 49–50, *Cohen*, 252 F. Supp. 3d. 140 (No. 1:16-cv-05158).

282. *Cohen*, 252 F. Supp. 3d at 158–61.

283. *Id.*

284. *Id.* at 157.

285. *Id.*

286. *Id.* at 161.

287. Force v. Facebook, Inc., 934 F.3d 53, 57 (2019).

288. *Id.* at 58–59, 65.

289. *Id.* at 64–65.

290. *Id.* at 71.

291. *Id.* at 64 ("In light of Congress's objectives, the Circuits are in general agreement that the text of Section 230(c)(1) should be construed broadly in favor of immunity.").

Facebook does not publish users' information and had no bearing on the plaintiffs' claims.[292] The court disagreed with the plaintiffs' contentions that the Facebook's algorithm turns Facebook into a publisher or a developer of the content because Facebook's algorithms are content neutral and merely display other users' content to users.[293] The court concluded that making Hamas's content more visible, available, and usable by using algorithms does not amount to developing content.[294]

Chief Judge Katzmann departed from the majority's conclusion on the issue of immunity for Facebook's suggestions for friends and content by its algorithms.[295] He explained that the sophisticated algorithms of Facebook bring users together after "collecting mountains of data" about their activity on and off its platform.[296] Chief Judge Katzmann reasoned:

> Facebook unleashes its algorithms to generate friend, groups, and event suggestions based on what it perceives to be the user's interests. If a user posts about a Hamas attack or searches for information on a Hamas leader, Facebook may "suggest" that the user become friends with Hamas terrorists on Facebook or join Hamas-related Facebook groups.[297]

Chief Judge Katzmann's opinion did not apply the immunity of § 230 for such functions, because, according to his judgment, it goes against the aim of § 230 to suppress indecent material.[298] When a plaintiff brings a claim that is not based on the content of information shown, but rather on the connections Facebook's

---

292. *Id.* at 70.

293. *Id.*

294. *Id.* ("[M]aking information more available is . . . an essential part of traditional *publishing*; it does not amount to 'developing' that information within the meaning of Section 230.").

295. *Id.* at 76 (Katzmann, C.J., concurring in part and dissenting in part).

296. *Id.* at 77.

297. *Id.* (citation omitted).

298. *See id.* at 79–80 ("The legislative history illustrates that in passing § 230 Congress was focused squarely on protecting minors from offensive online material . . . ."). *But see* Jeff Kosseff, *Correcting the Record on Section 230's Legislative History*, TECH. & MARKETING L. BLOG (Aug. 1, 2019), https://blog.ericgoldman.org/archives/2019/08/correcting-the-record-on-section-230s-legislative-history-guest-blog-post.htm [https://perma.cc/RQ2S-CAWK] (explaining that the broad language of Section 230's protection reflects the intent to protect the industry and its users' ability to communicate freely extends beyond prevention of indecent material).

algorithms make between individuals, the CDA does not bar relief.[299]

Chief Judge Katzmann concluded that Facebook may be immune under the CDA for allowing Hamas accounts, because "Facebook acts solely as the publisher of the Hamas users' content."[300] But the immunity does not apply when Facebook "conducts statistical analyses of that information and delivers a message based on those analyses."[301] Such activities in fact create networks of people, foment terrorism, and cause grave consequences.[302] Force appealed to the Supreme Court.[303]

Many other cases have focused on the presence of ISIS on social media. These cases have also been dismissed for lacking proximate cause under § 230. In *Fields v. Twitter, Inc.*,[304] the wife of Lloyd Fields, an American contractor who was killed by Abu Zaid in an ISIS shooting attack in Jordan,[305] contended that Twitter "knowingly permitted the terrorist group ISIS to use its social network as a tool for spreading extremist propaganda, raising funds and attracting new recruits," constituting "material support."[306] The district court dismissed the case, explaining that Twitter was immune under § 230.[307] Another ground for dismissal was the plaintiffs' failure to demonstrate that they were injured "by reason of" Twitter's conduct.[308] On appeal, the court of appeals ignored § 230, but affirmed the district

299. *See Cohen*, 934 F.3d at 80.

300. *Id.* at 83.

301. *Id.*

302. *See id.* at 86 (explaining that social media algorithms can be utilized by foreign governments to interfere in American elections, target emotions, and promote extremism and polarization).

303. Petition for a Writ of Certiorari, Force v. Facebook, Inc., No. 19-859 (U.S. Jan. 2, 2020); Mike Swift, *Facebook could face first Supreme Court challenge to Section 230 immunity*, MLEX (Jan. 13, 2020), https://mlexmarketinsight.com/insights-center/editors-picks/Data-Protection-Privacy-and-Security/north-america/facebook-could-face-first-supreme-court-challenge-to-section-230-immunity [https://perma.cc/9R6S-U76H].

304. 881 F.3d 739 (9th Cir. 2018).

305. *Id.* at 741–42.

306. Complaint at 1, Fields v. Twitter, Inc., 217 F. Supp. 3d 1116 (N.D. Cal. 2016) (No. 16-cv-00213).

307. *Fields*, 217 F. Supp. 3d at 1127–30.

308. This is the proximate cause argument. *See id.* at 1126–27.

court's decision on the ground that the plaintiffs failed to adequately plead proximate cause.[309]

A similar case, *Gonzalez v. Google, Inc.*,[310] followed the ISIS-driven terror attacks on La Belle Bistro and other coordinated attacks in Stade de France and the Bataclan Theater, which resulted in 130 deaths.[311] Nohemi Gonzalez, an America citizen, was killed in the La Belle Bistro and her family member filed a suit against Google (as owner of YouTube), Facebook, and Twitter.[312] The plaintiffs argued that the defendants knowingly permitted ISIS to use their social networks as a tool for spreading extremist propaganda, raising funds, and attracting new recruits in violation of the § 2333.[313] They further argued that media giants employed algorithms that promote terrorists' propaganda.[314] For example, Google's algorithms help users to locate similar videos and accounts, including videos and accounts related to ISIS even if they do not know the correct identifier.[315] Moreover, the plaintiffs alleged that Google derived revenues from ads and targeted ads to viewers based on algorithms that analyzed users and the videos they posted.[316] Relying on the *Roommates.com* case, the plaintiffs alleged that Google was a content creator.[317]

The court dismissed the case, distinguishing it from *Roommates.com* and concluding that Google did not materially contribute to the actual content of ISIS videos.[318] In addition, the court concluded the ads Google embedded next to ISIS content (which were themselves third party content), were not objectionable and did not play any role in making ISIS videos unlawful or encouraging individuals to commit acts of terror-

309. *Fields*, 881 F.3d at 741; *see also* Cain v. Twitter, Inc., No. 17-cv-02506-JD, 2018 WL 4657275, at *2 (N.D. Cal. Sept. 24, 2018) (dismissing the case because of the lack of proximate cause).

310. 282 F. Supp. 3d 1150 (N.D. Cal. 2017).

311. *Id.* at 1154.

312. *Id.* at 1153.

313. *Id.* The plaintiffs also argued that JASTA repealed the immunity provisions of the CDA, rendering section 230(c)(1) inapplicable in this case. *Id.* at 1157–61.

314. *Id.* at 1155.

315. *Id.*

316. *Id.*

317. *Id.* at 1168–69.

318. *Id.* at 1168–71.

ism.[319] Google used "neutral tools" in targeting ads and therefore did not develop the unlawful content.[320] An amended complaint in this case was also recently dismissed.[321]

Pennie v. Twitter, Inc.[322] involved a shooting of five Dallas police officers in 2016.[323] As with previous cases, the plaintiffs asserted that social media platforms allowed terrorists' content on their platforms and "developed" that content.[324] In addition, they argued that Google shared advertising revenues with FTO's and was not immune under § 230.[325] The court dismissed the case because of a lack of causal connection.[326] The court declined to resolve the question whether § 230 applied where the intermediary shared advertising revenues with users that had been designated as an FTO.[327] Recently, a second case regarding the Dallas shooting, with different plaintiffs, was dismissed in the court of Northern District of Texas also because of a lack of proximate cause.[328]

In *Crosby v. Twitter, Inc.*,[329] victims and families of deceased victims of the June 2016 mass shooting by Omar Mateen in an LGBT nightclub in Orlando filed a suit against the three media giants for providing a social media platform to terrorists.[330] They argued that videos and messages on social media radical-

---

319. *Id.* at 1168.

320. *See id.* ("Google's provision of neutral tools, including targeted advertising, does not equate to content development under section 230, because . . . the tools do not encourage the posting of unlawful or objectionable material.").

321. *See* Gonzalez v. Google, Inc., 335 F. Supp. 3d 1156, 1179 (N.D. Cal. 2018).

322. 281 F. Supp. 3d 874 (N.D. Cal. 2017).

323. *Id.* at 876.

324. *Id.* at 877, 891.

325. *Id.* at 884.

326. *Id.* at 892.

327. *Id.* at 891; *see also* Eric Goldman, *Fourth Judge Says Social Media Sites Aren't Liable for Supporting Terrorists—Pennie v. Twitter*, TECH. & MARKETING L. BLOG (Dec. 10, 2017), https://blog.ericgoldman.org/archives/2017/12/fourth-judge-says-social-media-sites-arent-liable-for-supporting-terrorists-pennie-v-twitter.htm [https://perma.cc/7ZNB-J3US].

328. Retana v. Twitter, Inc., No. 3:19-CV-0539-B, 2019 WL 6619218, at *1 (N.D. Tex. Dec. 5, 2019); *see also* Eric Goldman, *Twelfth Lawsuit Against Social Media Providers for "Materially Supporting Terrorists" Fails—Retana v. Twitter*, TECH. & MARKETING L. BLOG (Dec. 8, 2019), https://blog.ericgoldman.org/archives/2019/12/twelfth-lawsuit-against-social-media-providers-for-materially-supporting-terrorists-fails-retana-v-twitter.htm [https://perma.cc/77DX-PZ68].

329. 303 F. Supp. 3d 564 (E.D. Mich. 2018).

330. *Id.* at 565.

ized Mateen and triggered him to commit the terror attack.[331] The court based its decision on the merits of the substantive law,[332] and the lack of proximate cause, adopting a high threshold to meet this requirement.[333] The Sixth Circuit rejected the material support claim because of the proximate cause requirements.[334]

In *Clayborn v. Twitter, Inc.*,[335] the families of the victims of the 2015 attack in San Bernardino filed a suit against Facebook, Twitter, and Google.[336] They argued that in addition to providing the infrastructure for ISIS's activity, the defendants profited from ISIS by placing ads on ISIS postings.[337] Moreover, Google shared advertising revenues with ISIS.[338] The plaintiffs argued that by combining ISIS postings with advertisements, media giants create unique content.[339] The Northern District of California recently dismissed the case, concluding that the plaintiff failed to establish proximate cause between the social media platforms and the injuries.[340]

Sinclair v. Twitter, Inc.,[341] involved an ISIS terrorist attack in Barcelona on August 17, 2017, where a speeding truck was the

---

331. *Id.* at 569.

332. *Id.* at 573–74. The court reasoned that the plaintiffs did not prove that Mateen carried out the attack under ISIS's express direction. *Id.* at 573. In addition, the court concluded that there was no factual ground to suggest that the defendants "encouraged" Mateen to commit the attack or operated in concert with him. *Id.* at 574. The court also interpreted the knowledge requirement in the material support statute broadly and determined that the defendants lacked specific knowledge that they provided services to Mateen or that the defendants knew of any terrorist activities of ISIS that could be facilitated by the specific use of their services by any identified member or affiliate of ISIS. *See id.*

333. *See id.* at 579. The court reasoned that the plaintiffs did not prove that the defendants' furnishing of social media platforms to ISIS caused the specific attack. *Id.* Indeed, ISIS claimed responsibility for the attack after it was completed. *Id.* at 576. But nothing in the complaint hints of communication between ISIS and Mateen beforehand. *Id.* at 579.

334. Crosby v. Twitter, Inc., 921 F.3d 617, 624–26 (6th Cir. 2019).

335. Nos. 17-cv-06894-LB & 18-cv-00543-LB, 2018 WL 6839754 (N.D. Cal. Dec. 31, 2018).

336. *Id.* at *1–*3.

337. *Id.* at *6–*7.

338. *Id.* at *2.

339. *Id.* at *4.

340. *Id.* at *9. The Northern District of California later cited *Clayborn* in reaching the same conclusion. *See* Palmucci v. Twitter, Inc., No. 18-cv-03947-WHO, 2019 WL 1676079, at *3–*4 (N.D. Cal. Apr. 17, 2019).

341. No. C 17-5710 SBA (N.D. Cal. Mar. 20, 2019).

primary weapon; one of the victim's children filed an action against social media giants.[342] The court dismissed the suit because the allegations did not show proximate causation.[343] However, it left the door open for claims under wrongful death and "negligent infliction of emotional distress" state laws.[344]

In sum, most of the material support cases against media giants were dismissed based on § 230. Chief Judge Katzmann's minority opinion in *Force* presented a different approach[345] but the majority opinion, as well as many other courts, did not divert from the traditional interpretation of § 230. In other cases, the courts rejected the plaintiffs' arguments based on a narrow interpretation of the substantive law and a lack of a causal link between the intermediary's actions and the terrorist attack.

B. Normative Analysis

Imposing liability on intermediaries rests on a junction of several branches of law. It balances constitutional rights and public safety. It considers efficiency and cost-benefit analysis of legal obligations, and the technological context involves new questions and considerations of innovation policy. Providing a legal structure to identify values and outlining the right balance is a crucial judgment call, albeit a difficult one. The following Part focuses on three central situations that require nuanced examination: intermediation, failure to remove harmful content, and algorithmic targeting.

1. Freedom of Expression and Public Safety

Antiterrorism laws threaten freedom of speech, but terrorist groups threaten public safety.[346] How should democracies balance these two competing rights of freedom and safety? In the United States,[347] freedom of speech is more protected than in

342. *Id.* at 1.

343. *Id.* at 6–7.

344. *See id.* at 2, 14.

345. Force v. Facebook, Inc., 934 F.3d 53, 80–84 (2d Cir. 2019) (Katzmann, C.J., concurring in part and dissenting in part).

346. *See* Tsesis, *supra* note 17, at 617 ("[J]udges should protect constitutional rights of free speech, while recognizing congressional authority to enforce the Preamble of the Constitution's mandate to safeguard public safety.").

347. Oreste Pollicino & Marco Bassini, *Free speech, defamation and the limits to freedom of expression in the EU: a comparative analysis*, *in* RESEARCH HANDBOOK ON EU INTERNET LAW 508, 514 (Andrej Savin & Jan Trzaskowski eds., 2014).

other western democracies, whether it is political or commercial speech.[348] Courts and scholars have developed numerous theories about why free speech should receive special protection.[349] Freedom of speech promotes individual autonomy and self-fulfillment,[350] as well as the search for truth.[351] A free marketplace of ideas is essential for a liberal democracy.[352] Contemporary theories on democracy focus on protecting and promoting a democratic participatory culture.[353] Accordingly, freedom of speech is required to assure an individual's ability to participate in the production and distribution of culture.[354] This theory stresses both individual liberty and collective self-governance.[355]

The digital age, particularly the transition from the internet society to what Professor Balkin calls the "algorithmic society," push freedom of expression to the forefront, raising old concerns regarding expression.[356] The right balance must be struck between the benefits of free expression and the potential harm of inciting content to public safety. In the digital age, intermediaries host inciting content as they provide communication tools that enhance the flow of information.[357] They also target personalized recommendations on relevant content and con-

348. *See* Sorrell v. IMS Health Inc., 564 U.S. 552, 566 (2011); Jane R. Bambauer & Derek E. Bambauer, *Information Libertarianism*, 105 CALIF. L. REV. 335, 338 (2017); Tamara R. Piety, *"A Necessary Cost of Freedom"? The Incoherence of* Sorrell v. IMS, 64 ALA L. REV. 1, 4 (2012).

349. NEIL RICHARDS, INTELLECTUAL PRIVACY: RETHINKING CIVIL LIBERTIES IN THE DIGITAL AGE 10 (2015).

350. Joseph Raz, *Free Expression and Personal Identification*, 11 OXFORD J. LEGAL STUD. 303, 311–16 (1991) (arguing that freedom of expression enables self-determination of an individual by familiarizing public at large with his ways of life, allowing his preferences to gain public recognition and acceptability, and letting him know that he is not alone and his experiences are known to others).

351. *See, e.g.*, JOHN STUART MILL, ON LIBERTY 43–44 (Elizabeth Rapaport ed., Hackett Publ'g Co. 1978) (1859); JOHN MILTON, AEROPAGITICA: A SPEECH FOR THE LIBERTY OF UNLICENSED PRINTING 74–75 (T. Holt White ed., London, R. Hunter 1819) (1644).

352. *See* ALEXANDER MEIKLEJOHN, FREE SPEECH AND ITS RELATION TO SELF-GOVERNMENT 26 (1948).

353. Jack M. Balkin, *Digital Speech and Democratic Culture: A Theory of Freedom of Expression for the Information Society*, 79 N.Y.U. L. REV. 1, 3–4 (2004).

354. *Id.* at 4.

355. *Id.* at 1.

356. Balkin, *supra* note 148, at 1151–52.

357. *See supra* Part II.A.

nections.[358] This targeting may result in an enhanced flow of unlawful terrorist content and increased ease for terrorists to connect with each other and recruit.[359] Finally, intermediaries neglect to ban unlawful terrorists' content effectively and allow online terrorists' content to proliferate.[360] This neglect has consequences for public safety offline.[361] The law arguably should impose liability on intermediaries for hosting terrorist speech, targeting unlawful content, and neglecting to ban unlawful content consistently.

However, imposing liability on intermediaries for material support, or by any other regulatory means, may result in collateral censorship[362] because the new school of regulation affects the practical ability of users to speak.[363] It may result in censorship of legitimate speech, even if the intention is to remove unprotected speech.[364] Consequently, much content will be removed following referral units' requests from intermediaries, or even by using proactive algorithmic enforcement and over-blocking users' content and accounts without transparency.[365] This may happen even if the users are not affiliated with an FTO, because intermediaries tend to address context improperly and fail to distinguish terrorists' propaganda from other cultural content.[366]

---

358. *See supra* Part II.C.

359. *See id.*

360. *See supra* Part II.B.

361. *See id.*

362. Felix T. Wu, *Collateral Censorship and the Limits of Intermediary Immunity*, 87 NOTRE DAME L. REV. 293, 295–96 (2011) ("Collateral censorship occurs when a (private) intermediary suppresses the speech of others in order to avoid liability that otherwise might be imposed because of that speech.").

363. Balkin, *supra* note 143, at 2016.

364. *See* Ashcroft v. Free Speech Coal., 535 U.S. 234, 237 (2002) ("The overbreadth doctrine prohibits the Government from banning unprotected speech if a substantial amount of protected speech is prohibited or chilled in the process." (citing Broadrick v. Oklahoma, 413 U.S. 601, 612 (1973))). For more on censorship creep, see Citron, *supra* note 132, at 1049–71.

365. *See* Chang, *supra* note 181, 143–47.

366. *See* FRISCHMANN & SELINGER, *supra* note 64, at 146 ("[A]lgorithms can have a difficult time correctly identifying content that has context-specific meaning."); ZEYNEP TUFECKI, TWITTER AND TEAR GAS: THE POWER AND FRAGILITY OF NETWORKED PROTEST 150–51 (2017); Citron & Richards, *supra* note 134, at 1362 (explaining that algorithms are likely to result in over-removal of legitimate content because they are not sensitive enough to context). Tufecki provides an example in which "Facebook had adopted the U.S. State Department's list of 'terrorist organizations,' which included the Kurdish insurgent group, the PKK." *Id.* at 150. How-

Over-censorship is likely to hinder users' constitutional right to free speech on social networks.[367] It would curb users' ability to criticize the government and limit their ability to resist oppressive regimes.[368] It would probably infringe on speakers' autonomy, disrupt the exchange of ideas, and undermine civic and cultural participation. In addition, one might argue that by imposing liability on intermediaries, the government infringes on their right to free speech.

This chilling effect has a silver lining, and may be beneficial to some degree.[369] To strike the right balance between conflicting fundamental rights, courts and policymakers should focus on the role the intermediary plays in conveying the message, the severity of the message, and whether the message belongs to an unprotected category of "low-value" speech.[370] The role an intermediary plays in the offending speech should affect the preemptive measures taken against bearing liability, which, in turn, would influence the degree of censorship.[371] Furthermore, one should always keep an open eye on benefits of the service the intermediary offers to society as a whole.

For general platforms that do not encourage incitement in particular,[372] intermediation generally enhances freedom of ex-

---

ever, "Facebook fail[ed] to distinguish PKK propaganda from ordinary content that was merely about Kurds and their culture, or news about the group or the insurgency." *Id.* AI in content moderation can likely improve the accuracy of algorithmic enforcement. *See* Niva Elkin-Koren, *Fair Use By Design*, 64 UCLA L. REV. 1082, 1097 (2017). However, at this time, "[t]hese systems are . . . just not very good yet." *See* GILLESPIE, *supra* note 47, at 98.

367. Packingham v. North Carolina, 137 S. Ct. 1730, 1737 (2017) (noting that websites have become embedded in our culture as ways to communicate and exercise our constitutional rights).

368. TUFECKI, *supra* note 366, at 134.

369. *See* CASS R. SUNSTEIN, ON RUMORS: HOW FALSEHOODS SPREAD, WHY WE BELIEVE THEM, AND WHAT CAN BE DONE 74–75 (Princeton Univ. Press 2014).

370. *See* DANIELLE KEATS CITRON, HATE CRIMES IN CYBERSPACE 200 (2014); Genevieve Lakier, *The Invention of Low-Value Speech*, 128 HARV. L. REV. 2166, 2168 (2015); Geoffrey R. Stone, *Privacy, the First Amendment, and the Internet*, in THE OFFENSIVE INTERNET: PRIVACY, SPEECH, AND REPUTATION 174, 177 (Saul Levmore & Martha C. Nussbaum eds., 2010).

371. Tsesis, *supra* note 17, at 614 (arguing that computer intermediaries are not culpable for acting as instruments for third parties, even if the latter intended the material to be threatening).

372. *See supra* note 89. Platforms that are devoted to harmful expressions may infringe on the freedom of expression of victims and silence them. *See* Lavi, *supra* note 89, at 52–53. Thus, the infringement on the expression of victims may exceed the benefits of speech to speakers on these platforms. *See id.*

pression and does not aim to enhance terrorist content. Extending liability to these platforms for hosting terrorists' messages would result in collateral censorship.[373] Intermediaries would strive to reduce their risks by removing content automatically or proactively including protected speech without sensitivity to context.[374] Liability can also chill the development of communication tools and hinder users' ability to find relevant information and develop the marketplace of ideas.[375]

Sharing revenues with users is not intended to aid terrorism. It might be desirable if FTO's official profiles did not exist on social networks, but sharing revenues with users in general is not a serious threat to public safety, even if an FTO indirectly receives a small amount of money this way. Liability for sharing revenues with users may not have a serious effect on impeding users' speech today; it might result in the abandonment of this business model, which would reduce one of the incentives of users to speak.

Neglecting to remove terrorist content is different from limiting speech because of intermediaries' fear of liability. Policing harmful speech is subject to a strict scrutiny test.[376] But narrowly tailored liability could be limited to narrow categories of speech and specific methods of enforcement may reduce the chilling effect.

Certain forms of terrorists' speech on social media are unprotected by the First Amendment, such as promotion of "imminent lawless action,"[377] true intentional threats against individuals or groups,[378] or posts that seek to cooperate, legitimize, recruit, coordinate, or indoctrinate on behalf of groups listed on the State Department's list of designated terrorist organization.[379] The scope of these categories can be interpreted narrowly or broadly, influencing the degree of the chilling effect on

373. Citron, *supra* note 132, at 1039.

374. TUFECKI, *supra* note 366, at 160–62.

375. Seth F. Kreimer, *Censorship by Proxy: The First Amendment, Internet Intermediaries, and the Problem of the Weakest Link*, 155 U. PA. L. REV. 11, 16–17 (2006).

376. Turner Broad. Sys., Inc. v. FCC, 512 U.S. 622, 636 (1994).

377. *See* Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (per curiam); *see also* Tsesis, *supra* note 69, at 666–67 ("The incitement doctrine applies only to imminently dangerous statements and hence is of limited value to combat internet terrorist statements.").

378. Virginia v. Black, 538 U.S. 343, 359–60 (2003); Tsesis, *supra* note 69, at 670.

379. *See* Tsesis, *supra* note 69, at 670–75.

online speech.[380] In general, however, these forms of speech can be regulated without resulting in a conflict with free speech, because within the categories of unprotected speech, "the evil to be restricted so overwhelmingly outweighs the expressive interests, if any, at stake, that no process of case-by-case adjudication is required," and "the balance of competing interests is clearly struck."[381] Furthermore, unprotected speech arguably does not promote values of free speech because the right of free speech is granted to both sides. By imposing fear, inciting speech may chill the speech of others, hinder their autonomy, and compromise participation in the marketplace of ideas.[382]

As we turn to intermediaries' third party liability, the concern of collateral over-censorship arguably can be mitigated by restricting intermediaries' liability to instances where they are aware of the terrorist content. Extending intermediaries' liability to all content items on a platform would result in over-censorship, but imposing liability only for not taking down unprotected terrorists' content and FTO's accounts would lead to a desirable and proportionate chilling effect on speech. In such cases, the scope of liability is clearer and it does not require intermediaries to take proactive measures. Such a liability regime is superior to complete immunity, as immunity allows wild incitement, which results in great harm online and offline. Focusing on the type of speech and the intermediary's awareness strikes the right balance between free speech and public safety.

Algorithmic targeting might promote speech and enhance users' experiences as they meet like-minded people, but channeling users to specific content might create an echo chamber that limits the development of a free marketplace of ideas. An echo chamber might also strengthen terrorists' messages aimed at some users. Big data analysis and artificial intelligence (AI) might predict and modify users' behavior by utilizing their

380. *See, e.g.*, Guiora, *supra* note 71, at 140–41; Sherman, *supra* note 65, at 142; Leibowitz, *supra* note 28, at 818.

381. New York v. Ferber, 458 U.S. 747, 763–64 (1982).

382. Jeremy K. Kessler & David E. Pozen, *The Search for an Egalitarian First Amendment*, 118 COLUM. L. REV. 1953, 1994–95 (2018) ("[A]rguments involving *speech on both sides* focus on the degree to which one party's expressive activity compromises the ability of other private parties to exercise their own First Amendment rights.").

natural inclinations.[383] Algorithmic detection of a specific user that is interested in extreme ideas can result in targeting of recommendations, pushing users to consume unlawful terrorists' content and connect with members of FTOs, thereby changing the structure of the social network.

The ability of the intermediary to predict and influence users' behavior as a means to produce revenues raises a red flag. Intermediaries' liability for targeting can be justified to promote public safety. The chilling effect on recommendations is expected to be proportional because the intermediary can design the platform to avoid targeting unlawful content. It is true imposing liability in these cases may result in over censorship of legitimate recommendations, but intermediaries' self-censorship of recommendations is different than censoring users' speech because recommendations are based on third parties' content but are not the content itself. Arguably, the intermediary has more control over its algorithms relative to users' third party content and has more ability to avoid unlawful recommendations, especially in cases of defined forms of unprotected speech. Additionally, the potential harm for public safety that can result from recommendations on explicit incitement to terror is extensive and can bolster terror attacks outside the internet.

One may argue that imposing liability on intermediaries for algorithmic targeting of recommendations undermines their freedom to design platforms as they see fit. Imposing liability on targeting can also undermine intermediaries' freedom of expression.[384] Recommendations might not, however, be classified as speech, but rather a tool aimed to assist users in finding the right content.[385] On the other hand, recommendations arguably extend well beyond a functional tool. As have pointed out, the tool itself is an expression of the intermediaries' ideas[386] or

383. ZUBOFF, *supra* note 137, at 456–57.

384. *See* Tim Wu, *Machine Speech*, 161 U. PA. L. REV. 1495, 1515–21 (2013); *see also* Toni M. Massaro, Helen Norton & Margot E. Kaminski, *SIRI-ously 2.0: What Artificial Intelligence Reveals about the First Amendment*, 101 MINN. L. REV. 2481, 2483–84 (2017).

385. *See* Wu, *supra* note 384, at 1517–24 (differentiating speech and functional tools).

386. *See id.* at 1525. Professor Tim Wu refers to software navigation and map programs as harder cases of differentiation between communication of ideas and functionality. *Id.* He tends to believe that they are functional tools. *Id.* Other scholars adopt a broad approach to free speech and machine speech in particular. They

advice to users.[387] Assuming that recommendations are speech, intermediaries cannot have it both ways.[388] They cannot allege to be active speakers when seeking First Amendment protection and only navigation tools when facing tort liability. By enjoying the right of free speech, they undermine their immunity from civil liability as conduits and can bear liability as speakers.[389]

Furthermore, intermediaries collect mountains of data on users and use them to create algorithmic recommendations. This makes them powerful, knowledgeable speakers and justifies the application of a "listener-centered" approach for government regulation.[390] This approach permits regulation of speech for knowledgeable or powerful speakers when their expression frustrates the autonomy and self-governance of their listeners. Algorithmic recommendations that are directed at susceptible users and exploit their vulnerabilities to enhance the intermediaries' profits make the case for "listener-centered" approaches and justify imposing liability for targeting unprotected speech.

2. Corrective Justice

A central justification for imposing liability on intermediaries is corrective justice. Aristotelian philosophy defines corrective justice as a rectification of harm, wrongfully caused by one per-

---

argue that platforms direct users to material created by other and report it as newspapers report and thus, recommendations enjoy free speech protection. *See* Eugene Volokh & Donald M. Falk, *Google: First Amendment Protection for Search Engine Results*, 8 J.L. ECON. & POL'Y 883, 891 (2012). Another approach is that algorithms represent the message of its developers and are tied to human editorial judgement. *See* Stuart Minor Benjamin, *Algorithms and Speech*, 161 U. PA. L. REV. 1445, 1479 (2013).

387. *See* James Grimmelmann, *Speech Engines*, 98 MINN. L. REV. 868 , 874 (2014). In fact, this machine speech repeats users' speech and at times mimics it. Thus, this repetition promotes free speech. *See* Lavi, *supra* note 151, at 179–80.

388. *See* Oren Bracha & Frank Pasquale, *Federal Search Commission? Access, Fairness, and Accountability in the Law of Search*, 93 CORNELL L. REV. 1149, 1193 (2008). However, courts have reached different conclusions regarding search engines, recognizing intermediaries' right of free speech for page-rank and rejecting their liability for optimization. *See* Langdon v. Google, Inc., 474 F. Supp. 2d 622, 630 (D. Del. 2007); Search King, Inc. v. Google Tech. Inc., No. CIV-02-1457-M, 2003 WL 21464568, at *3 (W.D. Okla. May 27, 2003). These rulings have been criticized in literature. *See* Frank Pasquale, *Reforming the Law of Reputation*, 47 LOY. U. CHI. L.J. 515, 524–27 (2015); Wu, *supra* note 384, at 1496–1503, 1526–27 (describing the potential harm of computer-generated speech that invites regulation).

389. *See* RICHARDS, *supra* note 349, at 87.

390. *See* Helen Norton, *Powerful Speakers and Their Listeners*, 90 U. COLO. L. REV. 441, 443, 451 (2019).

son to another, by means of a direct transfer of resources from the injurer to the victim.[391] Accordingly, every interaction embodies correlative rights and duties that are imposed on both parties. This deontological, non-consequentialist concept focuses on bilateral interactions, which are not reliant on external values.[392]

Corrective justice theorists offer different motives for rectification—including conceptions of faults and rights[393]—based on responsibility,[394] and nonreciprocal risk.[395] Most theorists explain that there should be a causal link between the act and the consequence, but causation is not enough for imposing liability.[396] Negligence or moral fault must exist to justify compensation for the caused harm.[397]

The reason why harm is insufficient for justifying liability can be explained by nonreciprocal risks theory.[398] Liability exists when a person causes disproportionate risk, relative to the victim's risk-creating activity.[399] The entitlement to recover the loss is granted to all injured parties to the extent the risks imposed on them were nonreciprocal.[400] The goal is to distinguish between risk that violates individual interests and background risks that must be borne by society as a whole.[401]

---

391. 2 ARISTOTLE, *Nicomachean Ethics bk. V*, *in* THE COMPLETE WORKS OF ARISTOTLE 1781, 1786 (Jonathan Barnes ed., Princeton Univ. Press rev. Oxford trans. 1984).

392. *See* Ernest J. Weinrib, *Correlativity, Personality, and the Emerging Consensus on Corrective Justice*, 2 THEORETICAL INQUIRIES L. 107, 110 (2001).

393. *See* JULES L. COLEMAN, RISKS AND WRONGS 324–60 (1992); Benjamin C. Zipursky, *Civil Recourse, Not Corrective Justice*, 91 GEO. L.J. 695, 718 (2003).

394. Weinrib points out that tort doctrine constructs a tort relationship because liability treats the parties as doers and sufferers of the same injustice. *See* Weinrib, *supra* note 392, at 108–09; *see also* Stephen R. Perry, *The Moral Foundations of Tort Law*, 77 IOWA L. REV. 449, 449 (1992); Ariel Porat, *Questioning the Idea of Correlativity in Weinrib's Theory of Corrective Justice*, 2 THEORETICAL INQUIRIES L. 161, 169 (2001).

395. *See* George P. Fletcher, *Fairness and Utility in Tort Theory*, 85 HARV. L. REV. 537, 537–64 (1972).

396. *Id. But see* Richard A. Epstein, *A Theory of Strict Liability*, 2 J. LEGAL STUD. 151, 157–58 (1973) (arguing that harm itself is sufficient to justify compensation). However, this theory of strict liability, which focuses on factual causality, has been criticized. *See, e.g.*, Izhak Englard, *The System Builders: A Critical Appraisal of Modern American Tort Theory*, 9 J. LEGAL STUD. 27, 57–63 (1980).

397. Englard, *supra* note 396, at 57–63.

398. Fletcher, *supra* note 395, at 542–43.

399. *Id.*

400. *Id.*

401. *Id.*

In light of the bilateral correlative nature of torts, the literature on corrective justice tends to focus on "first order" liability of those who most directly and wrongfully caused an injury and not on "second order" liability of third parties that are not direct tortfeasors.

Intermediaries create the framework for terror attacks by allowing the activity and assisting it. Therefore, their actions are arguably more than a background risk and they can be liable for the consequences alongside the direct wrongdoer, because the corrective justice concept is also feasible when several wrongdoers caused the harm.[402] A counter argument might point out that the intermediaries did not cause the harm, and even when they bear culpability, there is no causal link between their activities and the harm terrorists' cause.

When an intermediary hosts content, provides communication tools, or shares revenues with users it does so for all types of content. It does not focus on terrorist content. These activities are merely a background risk. In such cases, the intermediary is not responsible for the harm caused to the victims and its liability is not justified according to corrective justice theory.

The case may be different when intermediaries fail to remove terrorist content that is unprotected by the First Amendment, such as fighting words, incitement to imminent lawless action, true threats or solicitations to commit crimes. In such cases, their liability is not the result of a pure omission, but instead their operation of the platform. The intermediaries are not mere bystanders. Arguably, if an intermediary acquires actual knowledge of a specific terrorist speech and fails to remove it and report the content to the authorities, it creates nonreciprocal risk that should not be immune to liability. Because moderation is an inherent role of twenty-first-century intermediaries, failure to remove upon knowledge extends beyond mere creation of a framework for harmful expressions. However, even if one accepts that failure to remove upon knowledge is a nonreciprocal risk, a causal link must exist to impose liability under corrective justice theory. This requirement of a causal link can

402. Richard W. Wright, *Allocating Liability Among Multiple Responsible Causes: A Principled Defense of Joint and Several Liability for Actual Harm and Risk Exposure*, 21 U.C. DAVIS L. REV. 1141, 1162 (1988).

be established only in extremely rare cases when the incitement is explicit and when a specific imminent terror attack is expected.[403]

When an intermediary personalizes content and targets users' unlawful content or suggests users to connect with a declared affiliate of an FTO, it bears direct responsibility and fault for the recommendation. In such cases, the intermediary allows a design of algorithmic recommendations that includes unlawful content,[404] such as incitement and postings that seek to recruit or coordinate on behalf of an FTO. But causal link requirements are met only between the intermediary and the algorithmic recommendations to users on inciting content, not between the recommendations and the terror attack. To justify liability for terrorist attacks, there should be evidence that the person who committed the attack was exposed to such recommendations and acted upon them. Otherwise, the intermediary may be responsible for the incitement,[405] or for contributing to the spread of inciting speech, though not for the terror attack itself.[406]

3. Efficiency

The perspective of efficiency focuses on the maximization of wealth and the efficient allocation of risks.[407] According to this perspective, legal rules aim to incentivize efficient conduct ex

403. In such cases, removal of the speech can be justified, even according the *Brandenburg* standard. Brandenburg v. Ohio, 395 U.S. 444, 447 (1969) (per curiam); Leibowitz, *supra* note 28, at 816.

404. Intermediaries can control the parameters at the base of the algorithms ex ante. On "policy neutral" vs. "policy directed" algorithms, see Tene & Polonetsky, *supra* note 151, at 137; Jack M. Balkin, *The Three Laws of Robotics in the Age of Big Data*, 78 OHIO ST. L.J. 1217, 1224 (2017). On government by design, see Deirdre K. Mulligan & Kenneth A. Bamberger, *Saving Governance-By-Design*, 106 CALIF. L. REV. 697, 701 (2018).

405. *See, e.g.*, Ryan Calo & Alex Rosenblat, *The Taking Economy: Uber, Information, and Power*, 117 COLUM. L. REV. 1623, 1684 (2017) (focusing on nontransparent manipulative practices in a related context of sharing platforms and suggesting that third parties' independent research can reveal some of these manipulative practices); Gerrard, *supra* note 38, at 4498 (showing inciting eating-disorder-related recommendations generated by intermediaries); Maayan Perel & Niva Elkin-Koren, *Black Box Tinkering: Beyond Disclosure in Algorithmic Enforcement*, 69 FLA. L. REV. 181, 202–05 (2017) (discussing algorithmic enforcement of copyright infringement).

406. This is because harm is directed at the general public, not at specific individuals. Therefore, legal civil action in tort law cannot be established.

407. See Richard A. Posner, *The Ethical and Political Basis of Efficiency Norm in Common Law Adjudication*, 8 HOFSTRA L. REV. 487, 488–91 (1980).

ante and promote welfare maximization ex post.[408] In this regard, courts should not consider the harm to victims in isolation. Rather they should include the benefits of an activity and any value that third parties gain from the activity. Their calculus should include all costs and benefits to society as a whole including the benefits of free speech and promotion of innovation.

Scholarly literature usually deals with the economic analysis of direct liability, but shies away from discussing third party liability.[409] However, based on the limited literature on this type of liability, I have argued elsewhere:

> [I]n some cases expanding liability to third parties is required when: (1) the enforcement of liability on the direct tortfeasor fails (for example, when the direct tortfeasor cannot be detected); (2) the third-party can monitor and control the direct wrongdoers; (3) sufficient incentives do not exist for private ordering and non-legal strategies; and (4) a legal rule can be applied at a reasonable cost.[410]

Pursuing a civil claim against the publisher of the incitement, or the direct terrorist attacker is possible and the law does not preclude civil remedies in cases of intentional criminal acts.[411] In the context of terrorism, however, there is a substantial risk of enforcement failure because the publisher may be anonymous or abroad and it would be difficult to bring him to comply with a judicial decision to compensate the victims.[412] The intermediary can be liable for the consequences alongside the direct attacker. Moreover, it might be difficult to collect com-

---

408. *See* J.R. Hicks, *The Foundations of Welfare Economics*, 49 ECON J. 696, 696 (1939); Nicholas Kaldor, Note, *Welfare Propositions of Economics and Interpersonal Comparisons of Utility*, 49 ECON. J. 549, 551 (1939).

409. Assaf Hamdani, *Gatekeeper Liability*, 77 S. CAL. L. REV. 53, 56–57 (2003) ("[T]he topic of third-party liability has received only scant attention by legal academics . . . . [L]ittle is known about the appropriate scope of third-party liability. Specifically, legal scholarship has little to say about the standard of liability that should apply to third parties.").

410. Lavi, *supra* note 48, at 882.

411. R.F.V. HEUSTON & R.S. CHAMBERS, SALMOND & HEUSTON ON THE LAW OF TORTS 227 (Sweet & Maxwell 18th ed. 1981). The intentional torts of assault and battery can give rise to both civil and criminal liability. Another example is civil compensation claims against rapists. *See* Julie Goldscheid, United States v. Morrison *and the Civil Rights Remedy of the Violence Against Women Act: A Civil Rights Law Struck Down in the Name of Federalism*, 86 CORNELL L. REV. 109, 113–15 (2000).

412. *See* Ronen Perry & Tal Z. Zarsky, *Liability for Online Anonymous Speech: Comparative and Economic Analyses*, 5 J. EUR. TORT L. 205, 237–38 (2014) (referring to a related context of intermediaries' liability to defamatory speech).

pensation from the direct attacker that may have died during the attack or cannot be found. In addition, the intermediary moderates the content and can control the speech published on the platform.[413] Private ordering is insufficient to tackle this problem.

Is it efficient to impose liability on intermediaries, or let the victims bear the costs?[414] To achieve efficiency, liability should be allocated to the cheapest cost avoider. Arguably, imposing liability on intermediaries is efficient and they are the cheapest cost avoiders of harm caused by terrorist content. They control the content on their platforms, make it easier to find it, and even encourage finding it.[415] This conclusion is valid even when they operate the platform and the recommendation system automatically through algorithms.[416] Restriction of recommendation systems and targeting is in use today. YouTube, for example, restricts its system to reduce harmful recommendations.[417] Likewise, Google announced that the company is planning to limit its advertisement targeting.[418] It is true that some

---

413. GILLESPIE, *supra* note 47, at 34.

414. *See* GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 68 (1970).

415. *See supra* Part II.A.

416. Intermediaries can control the parameters of the algorithms ex ante. *See* RONALD K.L. COLLINS & DAVID M. SKOVER, ROBOTICA: SPEECH RIGHTS AND ARTIFICIAL INTELLIGENCE 27 (2018) ("[Apple's] Siri has her limitations by design. She avoids controversy; she shuns opinions; she sidesteps medical, legal, or spiritual counsel; she eschews criminal advice; and she prefers the practice and factual to the ambiguous and evaluative."); Balkin, *supra* note 404, at 1233–36; Matthew U. Scherer, *Of Wild Beasts and Digital Analogues: The Legal Status of Autonomous Systems*, 19 NEV. L.J. 259, 280–90 (2018) [hereinafter *Wild Beasts*]; Matthew U. Scherer, *Regulating Artificial Intelligence Systems: Risks, Challenges, Competencies, and Strategies*, 29 HARV. J.L. & TECH. 353, 367 (2016) ("Even if that initial programming permits or encourages the AI to alter its objectives based on subsequent experiences, those alterations will occur in accordance with the dictates of the initial programming."); Tene & Polonetsky, *supra* note 151, at 137–42.

417. *See* YouTube Team, *Continuing our work to improve recommendations on YouTube*, YOUTUBE OFFICIAL BLOG (Jan. 25, 2019), https://youtube.googleblog.com/2019/01/continuing-our-work-to-improve.html [https://perma.cc/74KX-ZEPD] ("[W]e'll begin reducing recommendations of borderline content and content that could misinform users in harmful ways . . . .").

418. Tony Romm, *Google announces new political-ads policies that limit targeting but not all lies*, WASH. POST (Nov. 21, 2019, 6:41 PM), https://www.washingtonpost.com/technology/2019/11/20/google-announces-new-political-ads-policies-limiting-targeting-not-all-lies/ [https://perma.cc/UTV2-Y668]; Scott Spencer, *An update on our political ads policy*, GOOGLE CO. NEWS (Nov. 20, 2019), https://blog.google/technology/ads/update-our-political-ads-policy/ [https://perma.cc/3KJZ-EKX4].

technologies can lead to results that the intermediary cannot foresee ex ante.[419] But the intermediary can choose the technology it implements and limit it to a large extent beforehand.

Imposing liability on intermediaries would incentivize efficient moderation that mitigates the harm caused by terrorists' speech ex post.[420] It would reduce negligent design of recommendation systems ex ante and promote efficient deterrence. On the other hand, granting immunity to intermediaries incentivizes them to moderate irresponsibly, design unsafe recommendation systems, and externalize the damage caused to others. In addition, intermediaries normally have deeper pockets than individual victims and are better suited to reduce secondary costs by bearing the loss themselves or by spreading it to all their users.[421] An increase in litigation costs is expected, but imposing liability on intermediaries is better than the alternative of leaving the families of victims without a remedy.

An in-depth examination reveals that efficiency considerations fail to provide clear answers regarding the allocation of liability when considering overall market characteristics. Imposing liability on intermediaries will have little benefits in reducing radicalization and incitement because deactivating terrorists' accounts or removing their content will not prevent terrorists from reopening and republishing their content.[422] Thus, the intermediaries' efforts of removal may seem futile.

---

419. For example, the intermediary does not always foresee the exact results of the use of artificial intelligence and machine learning. *See* ADAM THIERER, ANDREA CASTILLO O'SULLIVAN & RAYMOND RUSSELL, MERCATUS CTR. GEO. MASON U., ARTIFICIAL INTELLIGENCE AND PUBLIC POLICY 31 (2017), https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf [https://perma.cc/X4JL-9DNU] ("[E]ven if the public *could* review them, the nature of machine-learning techniques can obviate the usefulness of review because the program is teaching itself.").

420. On this point, in the context of copyright infringement, see Douglas Lichtman & William Landes, *Indirect Liability for Copyright Infringement: An Economic Perspective*, 16 HARV. J.L. & TECH. 395, 398 (2003).

421. Perry & Zarsky, *supra* note 412, at 239.

422. *See, e.g.*, Corrected Complaint & Demand for Jury Trial at 136–41, Clayborn v. Twitter, Inc., 2018 WL 6839754 (N.D. Cal. Dec. 31, 2017) (No. 17-cv-06894-LB) ("According to the New York Times, the Twitter account of the pro-ISIS group Asawitiri Media has had 335 accounts. When its account @TurMedia333 was shut down, it started @TurMedia334 . . . . Below is a posting from Twitter captured on June 20, 2016. The individual is named 'DriftOne00146' and he proudly proclaims that this is the 146th version of his account. With only 11 tweets, this individual is followed by 349 followers. This is very suspicious activity.").

Intermediaries' effort to deactivate, or suspend terrorists' account and remove their content does not result in optimal enforcement. In fact, their activities can even backfire.[423] But these efforts are not completely futile because they increase the costs users must spend to find them.[424] In addition, to reduce the costs of enforcement, intermediaries are likely to adopt more efficient technology to identify FTOs' accounts and content.[425] Consequently, efficiency is likely to increase.

Another argument for not imposing liability on intermediaries is the risk that sanctions would distort access to digital markets and hinder positive externalities generated by intermediaries.[426] It could chill the development of communication tools and stifle innovative business models of revenue sharing.[427] Moreover, because asymmetry exists between the legal outcomes of false negative determinations of unlawful speech (liability) and exemption from liability for false positives, liability creates an incentive to remove more content than necessary for security.[428] Thus, it can lead to censorship of legitimate speech, chill recommendation systems, and hamper efficient design of platforms. Finding relevant information on the internet would be difficult,[429] and fewer innovative tools would be developed.[430] Imposing liability for not removing unlawful

---

423. SUNSTEIN, *supra* note 30, at 244 (explaining that suspending accounts can "create a more insular internal network that could promote an even more radicalizing force").

424. HARTZOG, *supra* note 94, at 223–26 (discussing in the context of online harassment).

425. *Id.* at 245; *see also* Citron & Richards, *supra* note 134, at 1362–63 ("Facebook employs algorithms to detect and remove terrorist speech. YouTube employs a tool called Content ID to prevent copyrighted material from being posted without the author's consent. The dominant online platforms—Twitter, Facebook, Microsoft, and YouTube—are developing an industry database that will collect hashes—or unique digital fingerprints—of banned violent extremist content for instant flagging, review, and removal.").

426. Kreimer, *supra* note 375, at 28–29.

427. *Id.*

428. *See* Assaf Hamdani, *Who's Liable for Cyberwrongs?*, 87 CORNELL L. REV. 901, 945 (2002) (discussing intellectual property infringements).

429. Seth Stern, Note, *Fair Housing and Online Free Speech Collide in* Fair Housing Council of San Fernando Valley v. Roommates.Com, LLC, 58 DEPAUL L. REV. 559, 590 (2009) ("If all websites strictly followed the Ninth Circuit's guidance, the Internet will eventually resemble a gigantic library with no cataloging system.").

430. The liability regime taxes innovation and influences its course. Evidence shows that innovation thrives under liberal liability regimes. *See, e.g.*, Kyle Graham, *Of Frightened Horses and Autonomous Vehicles: Tort Law and its Assimilation of Inno-*

speech is expected to disproportionately burden the rapid flow of information, free speech, and its positive externalities. As I have pointed out elsewhere, "Unlike traditional media, Internet content providers do not have the time or the resources to review and check every expression on their platform in real time."[431] Alternatively, algorithmic enforcement is not sensitive enough to context and may result in over-chill.[432]

When an intermediary moderates content, its liability for terrorists' content is secondary. Limiting liability for intermediaries to actual knowledge of unprotected speech and terrorists' accounts and subsequently not acting upon it can mitigate an over-chilling effect. Such a standard narrows uncertainty. It does not require proactive detection and moderation; it only requires intermediaries to remove specific categories of unprotected speech and official accounts designated as FTOs. As a result, removal is likely to focus on "low-value" speech. Liability concerns are less likely to lead to disproportionate over-censorship.[433] Limiting the scope of legal liability does not purport to preclude efficient efforts of moderation above this minimum standard. Thus, intermediaries may voluntarily remove wider categories of speech or develop proactive technologies for detection or removal.[434] Intermediaries might take these measures as a result of economic considerations and accountability, but not legal liability.

---

vations, 52 SANTA CLARA L. REV. 1241, 1251–52 (2012); Gideon Parchomovsky & Alex Stein, *Torts and Innovation*, 107 MICH. L. REV. 285, 314 (2008); Tal Z. Zarsky, *The Privacy-Innovation Conundrum*, 19 LEWIS & CLARK L. REV. 115, 126 (2015); Guy Pessach, *Deconstructing Disintermediation: A Skeptical Copyright Perspective*, 31 CARDOZO ARTS & ENT. L.J. 833, 864 (2013).

431. Lavi, *supra* note 48, at 883. Every minute 300 hours of videos are uploaded to YouTube. *See* Fred McConnell, *YouTube is 10 years old: the evolution of online video*, GUARDIAN (Feb. 13, 2015, 8:39 AM), https://www.theguardian.com/technology/2015/feb/13/youtube-10-years-old-evolution-of-online-video [https://perma.cc/F94P-GEHB].

432. *See* Citron, *supra* note 132, 1054–55.

433. *See* KENNETH A. BAMBERGER & DEIRDRE K. MULLIGAN, PRIVACY ON THE GROUND: DRIVING CORPORATE BEHAVIOR IN THE UNITED STATES AND EUROPE 242 (2015) (explaining that ambiguity regarding the exposure to liability leads businesses to adopt higher standards relative to the standards that would have been adopted under clear rules).

434. Klonick, *supra* note 112, at 1616–30 (explaining that intermediaries have created a voluntary system of self-regulation because they are aware of social and corporate responsibility and economically motivated to create a hospitable environment for their users to incentivize engagement).

When an intermediary personalizes recommendations on content or connections and targets users, the liability is not secondary because it directly spreads recommendations.[435] In such cases, the intermediary controls the design of the algorithm and can prevent recommendations that promote terror ex ante by programing algorithms that will not recommend content with inciting words or affiliates of FTO's as connections.[436] However, imposing liability on recommendations can cause a chilling effect on such systems, leading to less accurate algorithmic targeting or even to the elimination of these systems. Liability for recommendations may also stifle innovation.[437] Innovation may become too risky or expensive.[438]

Recommendations are an essential resource that reduce search costs and allow efficient engagement online. The imposition of liability, however, would probably cause a limited chilling effect on these systems so long as the liability focuses on unlawful recommendations. When the intermediaries' recommendations include inciting content, it is easier for terrorists to organize, recruit, and radicalize susceptible users, and it reduces terrorists' costs. Limiting unlawful recommendations on unprotected speech or official FTO connections is worthwhile even if insufficient sensitivity to context by algorithmic targeting reduces the accuracy of other recommendations. Furthermore, accurate recommendations, which intermediaries are economically incentivized to seek, would likely cause the development of more sensitive algorithms.[439]

---

435. Liability can be imposed even when someone repeats others. *See* Lavi, *supra* note 151, at 159.

436. On the ability to impose limitations on technology and learning algorithms in particular, see Scherer, *Wild Beasts*, *supra* note 416, at 280–90.

437. THIERER ET AL., *supra* note 419, at 36–37.

438. For example, machine learning might make it difficult for the intermediary to foresee results of their own algorithm because the algorithms learn and modify themselves through contacts with human users, the incorporation of obtainable data, or the insertion of new data. *See* Catherine Tremble, *Wild Westworld: Section 230 of the CDA and Social Networks' Use of Machine-Learning Algorithms*, 86 FORDHAM L. REV. 825, 837 (2017). Limiting the ability of algorithms and preventing this technology from including specific words in the design stage can result in less accurate recommendation. It may also discourage using innovative technologies as artificial intelligence.

439. Edmund Mokhtarian, *The Bot Legal Code: Developing a Legally Compliant Artificial Intelligence*, 21 VAND. J. ENT. & TECH. 145, 206 (2018) ("Rather than futilely attempt to micromanage these intelligent machines on an ad hoc basis, we likely

Imposing liability on intermediaries would have limited effect on innovation so long as liability remains neutral to technologies and does not depend on the adoption of a specific technology.[440] "Companies should generally have the freedom to design technologies how they please, so long as they stay within particular thresholds, satisfy certain basic requirements like security and accuracy, and remain accountable for deceptive, abusive, and dangerous design decisions."[441] Some innovators may "shy away from legally murky areas."[442] Nevertheless, there are other efficiency considerations to be balanced and "promoting innovation alone cannot be a sufficient justification for exempting intermediaries from the law."[443] There is an even more important reason why exemption from liability would be unwise. Overall immunity for all types of architecture designs "will yield a generation of technology that facilitates the behavior that our society has decided to prohibit."[444] Furthermore, exemption from liability may disincentivize intermediaries from developing safer and more efficient technologies.[445] Anyone who conducts business of any complexity must consult a lawyer about liability risks at some point. In many cases, innovation continues despite formidable legal regulations and ambiguity regarding the scope of liability. Furthermore, creativity and innovative thinking often thrive within

need to imbue them with the capability to comply with the legal systems in which they operate.").

440. *See* Nancy S. Kim, *Website Design and Liability*, 52 JURIMETRICS 383, 391–403 (2012); *see also* NATASHA DUARTE, EMMA LLANSO & ANNA LOUP, CTR. FOR DEMOCRACY & TECH., MIXED MESSAGES? THE LIMITS OF AUTOMATED SOCIAL MEDIA CONTENT ANALYSIS 6 (2017), https://cdt.org/wp-content/uploads/2017/11/Mixed-Messages-Paper.pdf [https://perma.cc/2BL5-UU85] ("Use of automated content analysis tools to detect or remove illegal content should never be mandated in law.").

441. HARTZOG, *supra* note 94, at 121.

442. Alex Kozinsky & Josh Goldfoot, *A Declaration of the Dependence of Cyberspace*, *in* THE NEXT DIGITAL DECADE: ESSAYS ON THE FUTURE OF THE INTERNET 169, 176 (Berin Szoka & Adam Marcus eds., 2010).

443. *Id.*

444. *Id.*

445. *See* Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345, 390 (2014).

constraint.[446] Thus, the concern about impeding innovation might be overstated.[447]

To sum up: imposing liability on intermediaries for terrorist attacks should not be ruled out. However, there are different roles that intermediaries fulfill and different types of speech. Therefore, a one-size-fits-all approach to intermediary liability is inappropriate.

IV. TAKING INFLUENCE SERIOUSLY

The broad reach of the internet and social media in particular has taken terror to another scale and level.[448] Terrorists' incitement, recruitment, and propaganda online result in terror attacks that pose a real threat to public safety and cause tremendous harm.[449] How should the law respond to this harm? Should online intermediaries that allow terrorist activities on their platforms and even contribute to them through content recommendations and targeting face liability? And if so, when? Normative analysis reveals that imposing liability on intermediaries for the results of terrorists' speech should not be ruled out altogether, but a more comprehensive framework is required. The following Part examines ways to overcome legal barriers in lawsuits grounded in material support that seek civil remedies for victims. Following this analysis, it offers using the "loss chances" doctrine and other possible legal tools that can lead to partial remedies and mitigate the problem of terrorists' incitement.

A. Overcoming Section 230's Barrier

Intermediaries are not mere conduits. As demonstrated in Part II, they provide communication tools, moderate content, and even influence speech by using algorithmic recommendation systems and other means. They can exacerbate or mitigate harm caused by illicit actors on their platforms.[450] However, the

446. *See* Joseph P. Fishman, *Creating Around Copyright*, 128 HARV. L. REV. 1333, 1336–37 (2015).

447. *See* Kozinsky & Goldfoot, *supra* note 442, at 176.

448. *See supra* Part I.B.

449. *See supra* Part III.A.

450. Klonick, *supra* note 112, at 1657–58; GILLESPIE, *supra* note 47, at 206–07.

current law provides immunity for intermediaries.[451] Thus, they are not treated as publishers of material they did not develop.[452] Courts have generally interpreted the immunity broadly.[453] However, this overall immunity scheme was constructed when the web was at its infancy.[454] As technologies advance and the web becomes more prevalent, the seriousness of terrorists' incitement increases and infringes on the public's sense of security and safety. Therefore, it is time to challenge the immunity regime and redefine it.

Recent scholarship acknowledges that twenty-first-century intermediaries structure, sort, and sometimes sell users' data. Thus, they cannot be treated as mere "passive conduits," and their role and duties should be reconceptualized.[455] Scholars have conceptualized intermediaries as governors and even advocated the imposition of public forum obligations on intermediaries, arguing that they should be treated as state actors. Such obligations would include holding intermediaries to standards of the First Amendment and requiring intermediaries complete content neutrality.[456] Other scholars have proposed viewing intermediaries as a hybrid of a conduit and media, recommending the imposition of some professional norms that apply to traditional media.[457] Recently, a new approach toward information fiduciaries analogizes intermediaries' duties towards

---

451. 47 U.S.C. § 230 (2018).

452. *See supra* Part III.A.2.

453. *Id.*

454. Leary, *supra* note 226, at 574; *see also* Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1952 (2019).

455. Kyle Langvardt, *Regulating Online Content Moderation*, 106 GEO. L.J. 1353, 1373 (2018).

456. K. Sabeel Rahman, *The New Utilities: Private Power, Social Infrastructure, and the Revival of the Public Utility Concept*, 39 CARDOZO L. REV. 1621, 1672 (2018). This position would impose public forum obligations on intermediaries. Such obligations are undesirable. Imposing public forum obligations will hinder efficient moderation and would do nothing to prevent third parties from using social media to manipulate end users. *See* JACK M. BALKIN, FIXING SOCIAL MEDIA'S GRAND BARGAIN 6 (Hoover Inst., Aegis Series Paper No. 1814, 2018), https://www.hoover.org/sites/default/files/research/docs/balkin_webreadypdf.pdf [https://perma.cc/G6YJ-UL2S]. Another difficulty in applying public law obligations "lies in the fact that internet platforms can 'evict' unwanted speakers without involving the courts." Langvardt, *supra* note 455, at 1367.

457. *See* GILLESPIE, *supra* note 47, at 43. This perspective supports the adoption of some of the professional norms of traditional journalism. *See* BALKIN, *supra* note 456, at 10. As Professor Balkin explains, however, the law still has a role to play. *Id.*

users' information with doctors and lawyers' fiduciary duties to their patients and clients.[458] The questions of the appropriate status of intermediaries and the scope of their general duties are beyond the purview of this Article. Be that as it may, the view that the overall immunity regime granted to intermediaries should adapt to the inflated influence online intermediaries have on users is gaining traction.[459]

Professor Danielle Keats Citron and Benjamin Wittes propose legislative changes to narrow down the scope of the immunity provision as a solution.[460] Under their proposal, the CDA's immunity provision would be available to operators only when they behave reasonably to stop illegal activity.[461] The consequence of that failure would not impose automatic liability, but rather remove the absolute shield from liability.[462] A continuous failure to remove an ISIS account despite repeated notifications might strip intermediaries' immunity.[463] This proposal is a good start, but it needs clearer standards regarding the unlawfulness of the content that will not enjoy immunity.[464]

A different approach allows the courts to discover the boundaries of immunity without a legislative change. Interme-

---

458. *See* Jack M. Balkin & Jonathan Zittrain, *A Grand Bargain to Make Tech Companies Trustworthy*, ATLANTIC (Oct. 3, 2016), https://www.theatlantic.com/technology/archive/2016/10/information-fiduciary/502346/ [https://perma.cc/7Q5B-RA4M]; *see also* Jack M. Balkin, *Information Fiduciaries and the First Amendment*, 49 U.C. DAVIS L. REV. 1183, 1226 (2016). The approach espoused by Professors Balkin and Zittrain would impose a duty on intermediaries to operate their platforms in good faith, with respect for users instead of manipulation. Balkin & Zittrain *supra*. Note that the information fiduciary approach raises issues regarding its feasibility, enforceability, and scope. For recent criticism that identifies tensions and ambiguities in the theory of information fiduciaries, as well as a number of reasons to doubt the theory's capacity to resolve them satisfactorily, see Lina M. Khan & David E. Pozen, *A Skeptical View of Information Fiduciaries*, 133 HARV. L. REV. 497 (2019).

459. *See, e.g.*, SYLVAIN, *supra* note 136, at 12 ("[T]hese developments undermine any notion that online intermediaries deserve immunity because they are mere conduits for, or passive publishers of, their users' expression.").

460. Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM. L. REV. 401, 418–19 (2017).

461. *Id.* at 419.

462. *Id.* at 420.

463. *Id.* at 418. This proposal allows media giants an exemption from liability for negligence. *See id.* It creates a type of notice and takedown regime. For an explicit proposal to adopt a notice and takedown regime, see Roter, *supra* note 101, at 1404–05.

464. *See* Citron, *supra* note 132, at 1052–55.

diaries structure, sort, target, and sometimes sell users' data.[465] By targeting content, the intermediary's algorithm not only repeats the content of users and advertisers, but also selects content for publication and displays different types of content to different audiences.[466] By doing so, the intermediary influences the context of the content and the magnitude ascribed to it. Therefore, intermediaries that design platforms and their code can be held responsible, at least in part, for creating or developing content.[467] This approach can be applied to algorithmic recommendations and targeting in particular.[468]

Stripping the immunity from co-development of content is in line with a broad interpretation of *Roommates.com*.[469] An intermediary's recommendations on content and connections are similar to the email mechanism in *Roommates.com*, which included only potential matches for roommates.[470] There are strong justifications for stripping the intermediary of immunity in these situations, especially if users did not positively articulate their preferences and the matching is a result of conclusions of algorithmic data processing. In contrast to the approach of the *Armslist* case, which referred to a design that can facilitate both lawful and unlawful activity, depending on the users' choice of use,[471] a recommendation system that includes unlawful recommendations is not a neutral tool. It exposes every user to different recommendations in light of algorithmic conclusions and is not based on users' positive choices. Thus, a

---

465. *See* Olivier Sylvain, *Intermediary Design Duties*, 50 CONN. L. REV. 203, 218 (2018).

466. *See id.*

467. *See id.* at 242; *see also* Tremble, *supra* note 438, at 866.

468. *See* KOSSEFF, *supra* note 223, at 188 ("As platforms increasingly develop more sophisticated algorithm-based technology to process user data it remains to see whether courts will conclude that they are responsible for the development of illegal content."). The dissenting opinion in *Force* reflected this approach. *See* Force v. Facebook, Inc., 934 F.3d 53, 76–77 (2d Cir. 2019) (Katzmann, C.J., concurring in part and dissenting in part).This opinion might be a step towards a change in courts' interpretation of § 230 in an algorithmic society. Chief Judge Katzmann focused on the function the defendant performed, referring to a decision in the context of commerce that imposed liability on Amazon based on the intermediary function. *Id.* at 81 (citing Oberdorf v. Amazon.com Inc., 930 F.3d 136, 153–54 (3d Cir. 2019), *vacated and reh'g granted by* 936 F.3d 182 (3d Cir. 2019)).

469. Fair Hous. Council of San Fernando Valley v. Roommates.com, LLC., 521 F.3d 1157, 1167–68 (9th Cir. 2008).

470. *Id.* at 1167.

471. Daniel v. Armslist, LLC, 926 N.W.2d 710, 722 (Wis. 2019).

specific susceptible user can receive only unlawful recommendations that can have profound influence on his decision to commit a terror attack.

Thus, § 230 of the CDA is not the main barrier for civil material support claims. Courts can strip intermediaries' immunity by interpretation even today. If the courts fail to narrow down this immunity, legislative changes can overcome this barrier.[472] These changes would strip immunity from intermediaries that fail to take terrorist content down upon learning of its existence and intermediaries that target unlawful content. Two questions remain: First, should the law impose obligations and liability on intermediaries for terrorists' incitement, after stripping immunity? And second when should the courts impose upon intermediaries an obligation to compensate victims' families for material support of terror?

B. Proximate Cause and Civil Remedies

To recover civil damages pursuant to the material support statutes, a plaintiff must establish that a defendant's conduct was the proximate cause of his injuries.[473] As explained earlier, some courts have outlined standards of substantial probability or foreseeable consequence of the specific act of support.[474] The thresholds of probability, however, were articulated in cases of donations or knowingly allowing the transfer of money to terrorist organizations, or directly assisting these acts.[475] In such cases, plaintiffs file an action against an entity that directly deals with a terror organization.[476] Consequently, there is an inherently direct connection between the defendants and the terror organization. In the context of social media, the courts have specifically articulated a requirement of proximity between the platform and the plaintiffs' injury. They have not settled for a lower threshold.[477] Social media is different from

472. *See* Citron & Wittes, *supra* note 460, at 418–19.

473. *See supra* Part III.A.2.c.

474. *See id.*

475. *See, e.g.*, *HLP*, 561 U.S. 1, 10 (2010).

476. The defendant could be, for example, a donor who directly donates money or a financial institution that allows an illegal transfer of money.

477. *See, e.g.*, Fields v. Twitter, Inc., 881 F.3d 739, 748–50 (9th Cir. 2018); Crosby v. Twitter, Inc., 303 F. Supp. 3d 564, 579 (E.D. Mich. 2018) ("To plead proximate cause, the plaintiffs must point to facts that tend to establish that (1) ISIS perpe-

donors, because unlike donors or transferors of money, social media companies operate platforms for all users to communicate, and there is no inherent direct connection between social media companies and terror organizations. This difference should lead to a different standard.

Intermediaries' liability for terror attacks cannot be fully analogous to general tort law cases on the duty to prevent crime.[478] For example, a therapist's duty of care to protect the intended victim of a dangerous patient is established only when the therapist has actual or constructive knowledge of the intentions of the patient and identification of the particular victim.[479] In contrast, intermediaries generally do not have knowledge about specific plans for an attack. A landlord's duty of care to protect tenants from foreseeable crimes committed on his premises[480] is also different from the case of intermediaries, because platforms host a tremendous amount of content and are different from premises in their characteristics. In addition, terror attacks are not committed on the platform itself. The intermediary can be responsible for allowing unlawful expressions on his platform, but it is difficult to predict which inciting speech was the trigger for the attack. The intermediary does not have a direct connection with terror organizations. Thus, it is difficult to establish a causal connection between failure to remove a specific post and a terror attack. This is the problem of latent harm that is difficult to match with precise wrong.[481]

The threshold of directness prevents victims' families from collecting full damages. There are strong policy considerations

---

trated the attack that injured them, and (2) the defendants' furnishing their social media platforms to ISIS caused the attack.").

478. *See* IZHAK ENGLARD, THE PHILOSOPHY OF TORT LAW 175–76 (1993).

479. *See* Tarasoff v. Regents of Univ. of Cal., 551 P.2d 334, 345 (Cal. 1976) (holding that mental health professionals have a duty to protect individuals who are being threatened with bodily harm by a patient); *see also* Gabe Maldoff & Omer Tene, *The Costs of Not Using Data: Balancing Privacy and the Perils of Inaction*, 15 J.L. ECON. & POL'Y 41, 64 (2019) ("Central to the ruling in *Tarasoff* was the fact that the professional could identify the particular victim.").

480. *See* ENGLARD, *supra* note 478, at 180; *see also* Kline v. 1500 Mass. Ave. Apartment Corp., 439 F.2d 477, 484 (1970) (adopting the foreseeability of criminal acts test).

481. *See* Omri Ben-Shahar, *Data Pollution*, 11 J. LEGAL ANALYSIS 104, 125 (2018) (giving a related example of liability for environment pollution, which "suffers from an acute problem of 'long tail'—latent harms that are difficult to causally match with precise wrongs").

running counter to waiving this threshold or replacing it with a lower one. In the absence of this threshold, intermediaries could be held responsible for all inciting speech published on their platforms that was neither removed nor reported before a terrorist's attack occurs. This scope of liability is too broad—almost limitless—and is not in line with normative considerations for liability.[482] Therefore, intermediaries' liability for full compensation of a terror attack's damages should be established only in extremely rare cases when the direct connection factor can be proven. Liability attaches to intermediaries only when terrorists' speech promotes an imminent lawless action[483] and the intermediary has actual knowledge of the speech, but fails to remove it and report it. For example, failure to act upon actual knowledge of a specific call to commit an act of terror that materializes could give rise to liability.[484] When a person other than the person that published the inciting content commits a terror attack, there should be a requirement to prove that the person who committed the attack was exposed to the inciting content that called for taking that specific lawless action.

Limiting compensation in such rare cases is in line with the material support doctrine. It may, however, result in underdeterrence because it fails to incentivize intermediaries to improve moderation and avoid unlawful targeting. It also leaves victims without any redress. Terrorists' speech on social media should be taken seriously. Therefore, policymakers should outline nuanced legal tools and measures to hold intermediaries accountable.

C. A New Framework of Intermediaries' Obligations Regarding Content, Algorithmic Targeting, and Terrorists' Accounts

Intermediaries can exacerbate or mitigate terrorists' speech, recruitment, and propaganda online.[485] Professor Citron and

---

482. *See supra* Part III.B. Full compensation is not in line with corrective justice and would lead to over-censorship and over-deterrence.

483. *See* Brandenburg v. Ohio, 395 U.S. 444, 447 (1969); In re White, No. 2:07cv342, 2013 WL 5295652, at *62–63 (E.D. Va. Sept. 13, 2013) (posting words on the internet alone was not sufficient evidence that the defendant's suggested actions were likely to be immediately carried out by his readers); *see also* Leibowitz, *supra* note 28, at 815–16.

484. Consider, for example, a post that contains the date and place of an intended terror attack.

485. *See supra* Part II.

Wittes propose that only intermediaries that behave reasonably to stop illegal activity should be immune to liability.[486] Failing to act against terrorist speech, however, can allow legal actions to proceed beyond preliminary stages.[487] This proposal leaves courts and policymakers to decide what constitutes illegal activity and what are reasonable steps to prevent it.

This Part proposes a defined legal duty of care regarding terrorists' unlawful content. Failure to meet the proposed standards of care would strip intermediaries of their immunity and allow the imposition of civil remedies, or even penal sanctions, against them. The proposed standards would reduce vagueness and allow intermediaries to manage their risks effectively. Unlike the overbroad standards proposed in the United Kingdom's white paper,[488] the proposed standard of care in this Article is tailored to unprotected speech and clearly defines the online harm. Thus, the proposal reduces the concern of undesirable political interference and the risk of disproportionate censorship.

The proposed framework would focus on moderation and algorithmic targeting. It would not impose special obligations on hosting, providing communication tools, and sharing revenues with users. In such cases, the intermediaries offer the same service to all users and do not prioritize terrorists' content over other providers' content. Thus, normative considerations do not advocate liability.[489]

1. Removal of Unprotected Speech Upon Knowledge

The first proposed change is narrowing immunity. Intermediaries should be exempt from liability only if they remove and report unprotected speech upon knowledge. A notice and

---

486. Citron & Wittes, *supra* note 460, at 419.

487. *Id.* at 420 ("Our proposal would not eliminate § 230's safe harbor. Instead, the safe harbor would be limited to providers or users that have taken reasonable steps to prevent or address the illegality of which plaintiffs are complaining.").

488. *See* SEC'Y OF STATE FOR DIG., CULTURE, MEDIA & SPORT & SEC'Y OF STATE FOR THE HOME DEP'T, ONLINE HARMS WHITE PAPER (2019), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/793360/Online_Harms_White_Paper.pdf [https://perma.cc/2CXX-M4D9] [hereinafter U.K. ONLINE HARMS WHITE PAPER].

489. *See supra* Part III.B. Imposing obligations and limitations regarding these functions would over-burden the flow of information and would result in inefficiency.

takedown regime is not new and governs the related context of copyright infringement.[490] Under this regime, the intermediary will not bear liability for terrorists' content if it expeditiously removes unprotected terrorists' speech and accounts of members of a designated FTO upon gaining actual knowledge of their existence. The intermediary can obtain this knowledge from private people who submit complaints, civil organizations, referral unit notifications,[491] or state authorities.[492] The intermediary should design clear mechanisms that make it easy to report unlawful content. If the intermediary opts to keep unprotected speech on its site, its action will not lead to automatic liability, but the intermediary will consequently lose immunity.[493]

Unprotected speech includes true threats,[494] fighting words, and social media postings that seek to cooperate, recruit, coordinate, incite, or indoctrinate users on behalf of designated terrorist organizations.[495] The removal obligation—as opposed to liability for full compensation—should not require imminence. Instead, a plaintiff would need to demonstrate that the speech directing, advocating, or encouraging lawless action caused a "substantial likelihood of a high level of harm."[496]

This proposal is thus narrowly tailored. Even if it indirectly leads to removal of protected speech, it is likely to pass the strict scrutiny test, as it places narrow limitations on speech

---

490. *See* 17 U.S.C. § 512 (2018).

491. *See* Chang, *supra* note 181.

492. The FBI and the Cybersecurity Unit at the Department of Justice are examples of such authorities.

493. On notice and takedown, see Perry & Zarsky, *supra* note 412, at 241–43. This regime does not impose liability on intermediaries that fail to actively monitor user-generated content published on their platforms.

494. *See* Tsesis, *supra* note 69, at 670.

495. *See* Raphael Cohen-Almagor, *The Role of Internet Intermediaries in Tackling Terrorism Online*, 86 FORDHAM L. REV. 425, 444 (2017); Tsesis, *supra* note 69, at 670–75.

496. Leibowitz, *supra* note 28, at 821. Some authors interpret the imminence test in *Brandenburg* broadly and reach similar results. *See* Tsesis, *supra* note 69, at 688–89 ("[W]here a person prods another on social media—such as Snapchat, WhatsApp, or Facebook Chat—to begin without delay a politically motivated attack, the statement can constitutionally, and should as a matter of social policy, be made actionable, if under the circumstances it is likely to incite such action."). Tsesis refers to the terrorists' speech and to the criminalization of identifiable terrorists' content. *See id.* at 697. The definition of unprotected speech is also applicable to removal obligations.

and mitigates the problem of disproportionate censorship.[497] Because of the broad influence of terrorists' activities on social networks, removing and reporting the worst speech is particularly important to public security.[498] This regime depends on reactive enforcement upon actual knowledge. Intermediaries can take additional measures of moderation, but they will not bear liability for failing to do so.

2. Safety by Design: Mitigating the Risk of Targeting of Unlawful Content and Recommendations

A second measure aims to limit the problem of targeting unlawful terrorists' content and recommendations by algorithmic designs and code.[499] Algorithmic recommendations of unlawful terrorists' speech can be described as "evil nudges,"[500] because they can push susceptible users to carry out a terror attack without forbidding any options or changing their economic incentives.[501] This practice raises a concern of incitement and manipulation by the intermediary.[502] The problem is exacerbated when the recommendations are personalized and target users who might have not actively searched for inciting content themselves.

---

497. *See* Tsesis, *supra* note 69, at 688–89. Even under a broad interpretation of imminence, in which the inclusion of direct calls for violence can be considered protected speech, the regulation can pass the strict scrutiny test because the states' interests are compelling and the notice and takedown regime is narrowly tailored.

498. Taking down inciting content is superior to reporting the content and leaving it online because not removing the content allows it to continue to spread and influence users to commit terrorist attacks. Indeed, the intermediaries should report to the authorities in cases of specific posts on upcoming incitement the same way they report on exploited children. *See* ROBERTS, *supra* note 107, at 106–07. Yet, The report, hoe should not replace the obligation to remove the inciting content. *See* Klein & Flinn, *supra* note 17, at 80.

499. *Supra* Part II.C.

500. *See* Lavi, *supra* note 89, at 1–2 (arguing that there should be legal liability for creating evil nudges that cause speech torts). Intermediaries' algorithms recommend unlawful content unwittingly, without an intent to incite. However, these recommendations are not arbitrary and are instead designed to target susceptible users to the type of recommended content. Thus, algorithmic design can mitigate the practice of unlawful algorithmic recommendations.

501. *Id.* at 1. For more on nudges, see RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008).

502. *See* VAIDHYANATHAN, *supra* note 138, at 151–52 (raising questions about what a democracy would look like if Facebook's algorithm governed the art of science and persuasion).

To tackle this problem and enhance safety, this Part proposes the concept of "safety by design." This type of regulation, first identified by Professor Lawrence Lessig,[503] proposes technology-based solutions for preventing harm inflicted by the flow of information. Engineering decisions can unleash new technology not previously contemplated by the law and affect fundamental rights. Scholars and policymakers have already explored the influence of technological design and its potential to infringe on values or promote a variety of values in the design stage.[504] Ethics alone cannot solve the problem of algorithmic incitement and there should be legal constraints on how algorithms are used.[505] This concept may be used on the architecture of the platforms and on algorithmic code.[506]

An intermediary that targets unlawful content that incites to terror and recommends unlawful connections of FTO members should not enjoy complete immunity because recommendation systems are high-risk automated systems; the intermediary proposes recommendations by itself and arguably provides or at least develops the content by taking it out of its original context and targeting vulnerable users.[507] Furthermore, the intermediary directly manipulates users to behave unlawfully.[508] This conclusion remains true even if the targeting is committed

---

503. *See* LAWRENCE LESSIG, CODE: VERSION 2.0, at 123 (2006) (identifying four key forces that regulate an online environment: "the law, social norms, the market, and architecture").

504. Designers and even lawmakers can protect values of privacy by design. "Privacy by design" is an approach that incorporates thinking about privacy-protective features and implementing them as early as possible. *See* BAMBERGER & MULLIGAN, *supra* note 433, at 32, 178 (2015). Regulators have discovered the benefits of using design to protect privacy, put forth guidelines, and incentivize stakeholders to adopt this approach as part of their business models. *See* CHRIS JAY HOOFNAGLE, FEDERAL TRADE COMMISSION PRIVACY LAW AND POLICY 190–92 (2016); *see also* Mulligan & Bamberger, *supra* note 404, at 701. Value-sensitive design is an approach that advocates identifying human needs and values and taking them into account in the design process. *See* Noëmi Manders-Huits & Jeroen van den Hoven, *The Need for a Value-Sensitive Design of Communication Infrastructures*, *in* EVALUATING NEW TECHNOLOGIES 51, 54 (Paul Sollie & Marcus Düwell eds., 2009); Deirdre K. Mulligan & Jenifer King, *Bridging the Gap Between Privacy and Design*, 14 U. PA. J. CONST. L. 989, 1019 (2012).

505. *See* Daniel Susser, *Ethics Alone Can't Fix Big Tech*, SLATE (Apr. 17, 2019, 11:45 AM), https://slate.com/technology/2019/04/ethics-board-google-ai.html [https://perma.cc/M5WK-K27P].

506. Levy & Barocas, *supra* note 143, at 1230.

507. *See* Lavi, *supra* note 151, at 195.

508. Sylvain, *supra* note 465, at 275; Tremble, *supra* note 438, at 866.

automatically, because algorithmic, independent decisions are constrained by ex ante data choices and instructions.[509] As for the scope of their liability, intermediaries are free to design their technology as they see fit, but they should be subject to basic requirements to keep users safe, and they should have a duty of care to internalize the cost they impose on society through algorithmic recommendation and targeting of unprotected content.[510] This duty of care focuses on the relations between the user and the intermediary. It can be applied to algorithmic targeting[511] and protects security in society in general. This duty creates obligations to avoid targeting unprotected speech that directs, advocates, or encourages lawless action and incites and manipulates susceptible users to engage in terrorism. To meet this duty, intermediaries should utilize the concept of safety by design and instruct code developers to limit their code. Intermediaries can impose a barrier on the recommendations ex ante and avoid algorithmic recommendations with specific words or connections.[512] Limitations by design are applied to other technologies and can be transplanted to the context of algorithmic recommendations as well.[513]

Indeed, this solution may lead intermediaries and code developers to limit the boundaries of learning algorithms in the design stage of the code, which may result in less accurate recommendations. In the alternative, however, imposing a duty of care may incentivize intermediaries to develop more accurate technology and algorithms that will achieve both efficiency and accuracy.[514] Even if this scenario would not fully materialize, intermediaries cannot enjoy the rights of free speech and freedom to design without responsibility.[515] Because algorithmic incitement has power to influence users, promote terrorism, and cause tremendous harm, intermediaries should avoid targeting unprotected speech and connections, such as recommending connecting with members of an FTO.

---

509. That the recommendations are automatic does not change the conclusion, because the intermediary can impose barriers on the algorithm ex ante. *See* Balkin, *supra* note 404, at 1224; Mulligan & Bamberger, *supra* note 404, at 701.

510. HARTZOG, *supra* note 94, at 126.

511. *See* Balkin, *supra* note 404, at 1224.

512. *See id.*

513. COLLINS & SKOVER, *supra* note 416, at 27.

514. Mokhtarian, *supra* note 439, at 179.

515. *See* HARTZOG, *supra* note 94, at 121–26; RICHARDS, *supra* note 349, at 87.

Because recommendations are personalized, it is difficult to discover infringement of the obligation to avoid unlawfulness; most platforms avoid disclosing operational details on content recommendation practices.[516] However, researchers, members of civil rights organizations, and users can discover targeting of unlawful content in some instances.[517] Moreover, regulators can call upon or even fund independent researchers specifically to analyze digital practices to uncover inciting algorithmic systems of platforms.[518] In addition, policymakers can encourage challenging nontransparent recommendation systems by using a proactive method of "black box tinkering."[519] This method encourages public activism and engagement in checking the practices of automatic enforcement systems.[520] As a result, intermediaries could be held responsible for these nontransparent practices.[521]

Disclosure by inside employees that might be motivated by a concern for others' wellbeing, and who can shed light on the algorithms, may be another way to improve flawed practices and accountability. In a related context, Professor Sonia Katyal proposed encouraging greater transparency in algorithmic practices by adopting whistleblower protections.[522] This solution might be applicable to inciting policy-directed targeting. Protecting individual employees of media giants who come forward to address issues of flawed practices of targeting would create incentives to disclose information and may enable greater mitigation of harm and improved accountability.

A more comprehensive approach for promoting algorithmic safety and accountability is to develop new frameworks and methods of algorithmic oversight and public regulation.[523] The

---

516. COHEN, *supra* note 165, at 136.

517. *See, e.g.*, Gerrard, *supra* note 38, at 4504 (discovered algorithmic targeting of "harm to self" content).

518. *See* Calo & Rosenblat, *supra* note 405, at 1684.

519. *See* Perel & Elkin-Koren, *supra* note 405, at 198–211 (arguing that public engagement in checking the practices of automatic enforcement systems can mitigate the problem and enhance awareness of biased algorithms).

520. *See id.*

521. *See id.*

522. Sonia K. Katyal, *Private Accountability in the Age of Artificial Intelligence*, 66 UCLA L. REV. 54, 126 (2019).

523. COHEN, *supra* note 165, at 267 ("The task of ensuring progress towards broadly distributed development , sustainability and algorithmic accountability is not one of courts alone, or even primarily; it will also require new methods of

Article will touch upon such methods and address them in Part IV.C.4.c.

3. Safe Haven: Outlining a Lenient Liability Regime for Adopting Safety by Design, Best Practices, and Monitoring

What should the legal scope of liability for algorithmic targeting be? Should the law sanction unprotected recommendations even if the intermediary did not intend to select unlawful recommendations, or should it settle with voluntary measures of prevention? The private industry has an important role in promoting algorithmic accountability for safer algorithmic targeting.[524] However, there are different technology and media companies with different business models and agendas. In addition, companies outsource responsibility to fulfill legal requirements and duties to engineers at third party technology vendors. They see the obligations through a corporate lens that aims to maximize their profits, rather than through a substantive lens.[525] Although voluntary regulation is highly important, relying on it alone is insufficient. Therefore, the starting point is subjecting intermediaries to a legal duty of care to avoid targeting unprotected speech and recommendations. Failing to comply with this duty can result in legal liability.[526]

In many cases the algorithm is policy neutral,[527] and specifically targeting unlawful content is an unwitting consequence that the intermediary might not have aimed for. As explained in the last Part, liability might over-chill speech and reduce the

---

administrative oversight and new thinking about the approach relationship(s) between administrators and courts").

524. Katyal, *supra* note 522, at 61 (arguing that because the state alone cannot solve the problem of algorithmic accountability, a solution should involve the private industry).

525. *See* Ari Ezra Waldman, *Privacy's Law of Design*, 9 U.C. IRVINE L. REV. 1239, 1245–59 (2019).

526. Liability should be governed by the concept of product liability, even if the default regime is negligence. *See* Katyal, *supra* note 522, at 126; Waldman, *supra* note 525, at 1263–66. Even intermediaries that use learning algorithms should be aware of the risks of harm and can make sure that the designers impose limitations on the systems' culpabilities. *See* Scherer, *Wild Beasts*, *supra* note 416, at 280–90. As an alternative to the strict liability regime which is likely to over-chill free speech, liability can be imposed according to a standard of "the reasonable algorithm." *See* Ryan Abbott, *The Reasonable Computer: Disrupting the Paradigm of Tort Liability*, 86 GEO. WASH. L. REV. 1, 5–6 (2018).

527. *See* Tene & Polonetsky, *supra* note 151, at 132.

accuracy of legitimate recommendations. This result reduces fairness and efficiency, especially under a rigorous standard of strict liability, but also under an ambiguous negligence standard. On the other hand, even an algorithm that is policy neutral can pose a great risk by targeting inciting content to susceptible recipients and pushing them to commit violent terror attacks. An overall immunity regime creates disincentives for intermediaries to avoid targeting unlawful recommendations and result in under-deterrence and inefficient levels of risks to the public's safety. To balance efficiency, fairness, and public safety, a safe haven regime for algorithmic targeting should be promulgated. Under this regime, liability would not be imposed for failing to provide perfect safety. Instead, this regime aims to incentivize all companies to comply with specific requirements and minimize disproportionate risk of unlawful recommendation.

The starting point for targeting unprotected speech is a negligence theory of liability. Intermediaries that choose to opt in to a safe haven program can gain certainty regarding the scope of liability. The proposed safe haven includes concrete obligations and duties of care. It can apply only to general purpose platforms and not to purely ideological platforms that are devoted to incitement to terror and hate speech without legitimate purpose.[528] The safe haven requirements will focus on minimizing risk for unprotected speech as a minimum standard of care. Indeed, intermediaries are encouraged to reduce the risk of extremist recommendations and to reduce the visibility of violent content. However, this extra level of care is voluntary.

By complying with the requirements of the safe haven program, intermediaries will be exempt from civil or criminal fault-based liability. Liability will be limited only to knowledge-based targeting of unprotected speech and recommendation, or for knowingly failing to fix code that causes inciting algorithmic recommendations. Intermediaries can still bear liability if civil society organizations, private people, or monitoring systems report recommendations of inciting con-

528. Platforms that are focal points for incitement, such as Gab, cannot benefit from a safe haven, and their liability will be governed under a standard of negligence, or even under a standard of inducement. *See* Lavi, *supra* note 89, at 5.

tent, or if the intermediary received notifications yet continues to exhibit such recommendations.[529]

The safe haven program will require intermediaries to involve attorneys in the design process and work with governmental authorities to implement and comply with safety standards in algorithmic recommendation systems.[530] First, it will require intermediaries and industry bodies to develop and adopt safety technologies and best practices and apply them in algorithmic design.[531] Standards of safety by design aspire to prevent automatic suggestions of terrorist content and reduce the risk for unlawful recommendation of content. Industry and government experts would review the standards every year and update them as technology develops.

Second, recommendation systems can involve artificial intelligence algorithms that can learn and operate unexpectedly. However, that the system may operate in unpredictable ways is predictable. It is inefficient and unjust to provide a safe haven for knowingly operating an unpredictable system without minimizing its risks. Therefore, intermediaries that aim to comply with the safe haven requirements should limit their operations and disable the ability to create unlawful recommendations ex ante at the design stage.[532] Alternatively, a solution of monitoring systems ex post can mitigate the risk of diverting from initial programing. Automatic monitoring systems would review the algorithms, notify the intermediary that

529. *See* Kim, *supra* note 440, at 416–18.

530. *See* Waldman, *supra* note 525, at 1283 (in the related context of privacy by design).

531. This regulatory concept was recently found in the regulatory framework for accommodating terrorist online harm in the United Kingdom that includes a risk based duty of care in algorithmic selection of content. *See* U.K. ONLINE HARMS WHITE PAPER, *supra* note 488, at 70, 72 ("[C]ompanies should take [reasonable steps] to ensure that their services are safe by design . . . . Companies will be required to ensure that algorithms selecting content do not skew towards extreme and unreliable material in the pursuit of sustained user engagement.").

532. Limiting the function of the system to avoid specific topics at the design stage is possible even when AI is involved. Apple's Siri demonstrates this point. *See* COLLINS & SKOVER, *supra* note 416, at 27; *see also* YouTube Team, *supra* note 417 (discussing the practice of limiting harmful recommendations applied by YouTube).

it should correct algorithmic failure, and reprogram or redirect the algorithmic recommendation system ex post.[533]

Third, intermediaries should implement reporting systems that allow private people and civil society organizations to report unlawful recommendations on inciting speech efficiently and allow for correction.

Fourth, intermediaries should submit transparency reports regarding the technological mechanisms to regulators. These reports would explain how their algorithms operate and select content, thereby reducing the risk for unlawful algorithmic recommendations.[534] Transparency obligations will enable regulators to ensure that the intermediary complies with the safe harbor obligations and to sanction noncompliance.[535]

4. Remedies, Sanctions and Regulatory Tools

a. Tort Law: Loss of Chances Doctrine

This Article explains that the proximate cause requirement makes it difficult to establish liability on online intermediaries for terror attacks after a failure to remove specific speech or a failure to prevent a specific content recommendation.[536] However, there is a good reason to believe that inciting speech on social networks inspires terrorism.[537] An intermediary that knowingly fails to remove unprotected terrorist speech or de-

533. On ex post monitoring systems as part of a liability regime of robotic functions, see Omri Rachum-Twaig, *Whose Robot is it Anyway?: Liability for Artificial-Intelligence-Based Robots*, 2020 U. ILL. L. REV. (forthcoming 2020) (manuscript at 33).

534. Algorithmic code is usually a trade secret. However, the requirement is not to expose the code but rather to explain it. For a similar proposal, see U.K. ONLINE HARMS WHITE PAPER, *supra* note 488, at 45. In addition, transparency that is limited only to the regulator should not be ruled out. *See* Tal Z. Zarsky, *Transparent Predictions*, 2013 U. ILL. L. REV. 1503, 1540; *see also* Hannah Bloch-Wehba, *Access to Algorithm*, 88 FORDHAM L. REV. (forthcoming 2020) (manuscript at 49) ("Faced with demands for more transparency, courts and litigants have sometimes reached an apparent compromise: protective orders, coupled with nondisclosure orders, that permit disclosure to the parties while preventing disclosure to the general public.").

535. Algorithmic oversight should improve and extend beyond transparency obligation as policymakers and regulators develop and adopt more substantive methods for algorithmic evaluation and public regulation. On a more comprehensive frameworks for algorithmic evaluation and public regulation, see the proposals for public regulation and algorithmic impact assessment in Part IV.C.4.c.

536. *See supra* Part IV.B.

537. *See supra* Part I.B.

signs an algorithmic recommendation system that targets recommendations for inciting content increases the risk for terrorist attacks. Victims of terror attacks and their families have sustained harm that might have been caused by the intermediary's behavior. The casual connection, however, is an uncertain factor. Namely, it is unclear whether the defendant's wrongdoing actually violated the plaintiffs' protected interest.[538] Many scholars argue that in cases of systematic infliction of harm and uncertainty in causation, an adherence to an all-or-nothing solution is inappropriate.[539] Therefore, plaintiffs should be able to recover their loss under the lost chances doctrine. The compensation would be proportionate to the probability of loss of chances even if the loss of chances is below fifty percent.[540] The law has adopted the lost chance doctrine in different jurisdictions mainly in the fields of medical malpractice and mass torts.[541]

Indeed, U.S. courts are inconsistent in applying this doctrine.[542] However, the loss chances doctrine may provide a solution for the problem of uncertain causation in intermediaries' liability for terrorists' content. Imposing proportional liability on the intermediary is justified from a corrective justice perspective. It imposes compensation on intermediaries according to the actual damage they caused and allows terrorist victims and their families to get partial compensation. Applying the

538. These factors define uncertainty depending on whether the defendant violated the plaintiff's protected interest. *See* ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 125–29 (2001); Marc Stauch, *Causation, Risk, and Loss of Chance in Medical Negligence*, 17 OXFORD J. LEGAL STUD. 205, 223 (1997).

539. *See, e.g.*, PORAT & STEIN, *supra* note 538, at 125–29.

540. *See id.* at 127.

541. In these situations, the plaintiff asserts that a certain percentage of his chances of recovery were lost as a result of the defendant's negligent omissions. *See* Benjamin Shmueli, *"I'm Not Half the Man I Used to Be": Exposure to Risk Without Bodily Harm in Anglo-American and Israeli Law*, 27 EMORY INT'L L. REV. 987, 998 (2013).

542. *Compare* Herskovits v. Group Health Coop. of Puget Sound, 664 P.2d 474, 476–77 (Wash. 1983) (adopting this approach) *with* Cooper v. Sisters of Charity of Cincinnati, Inc., 272 N.E.2d 242 97, 104 (1971) *and* Hiser v. Randolph, 617 P.2d 774, 779 (Ariz. Ct. App. 1980) (rejecting this approach). Although there is no consensus for applying this doctrine, courts are more willing to adopt it relative to the increased risk doctrine that mirrors it. *See* Shmueli, *supra* note 541, at 998; *see also* Daniel J. Solove & Danielle Keats Citron, *Risk and Anxiety: A Theory of Data-Breach Harms*, 96 TEX. L. REV. 737, 740 (2018) (proposing to apply the increased risk doctrine on data breach cases and referring to anxiety risk as actual harm).

doctrine is also justified from an efficiency perspective. It results in optimal compensation and solves the under-deterrence problem that would have been the result otherwise.[543]

It should be noted that the loss chances doctrine aims to compensate for harm that already occurred. It is different from the "increased risk" doctrine that aims at compensation for increased risk for future harm that plaintiffs might seek to apply and was at the base of the plaintiffs' suit in *Cohen v. Facebook, Inc*.[544] Applying the doctrine of loss of chance is more feasible than applying the increased risk doctrine on future attacks.[545] In contrast to the increased risk doctrine, loss chance doctrine deals with actual harm that already occurred. Actual specific victims may receive partial compensation from the worst actors that have knowingly failed to remove unprotected inciting content or recommendations for unprotected speech upon notice.

b. Criminal Prosecution

Criminal law allows the Justice Department to file criminal suits against companies for violating the true threats,[546] or the material support statutes.[547] Criminal liability could include monetary fines or takedown orders against terrorists' accounts.[548] A court might also issue an injunction to deploy software for taking down unprotected terrorists' expressions or

---

543. *See* PORAT & STEIN, *supra* note 538, at 128–29.

544. In *Cohen*, 20,000 people filed an action and argued that future attacks threaten them. Cohen v. Facebook, Inc., 252 F. Supp. 3d 140, 145–46 (E.D.N.Y. 2017). Compensating for future attacks is less desirable in such cases. Unlike data breach cases which include a limited group of people whose personal data has been breached, there is no defined group of plaintiffs, and allocating compensation is thus problematic. *See* Ben-Shahar, *supra* note 481, at 24; Solove & Citron, *supra* note 542. In the case of increased risk doctrine—as opposed to the lost chance doctrine—the uncertainty over causation is not the only problem. There is also uncertainty regarding who will be the actual victims.

545. Courts are not likely to recognize fear of future terror attacks as actual harm, when physical harm had not yet occurred. In *Spokeo, Inc. v. Robins*, 136 S. Ct. 1540 (2016), the Supreme Court held that harms need not immediately translate into an injury if there is a significant risk of a real harm occurring later. *See id.* at 1549. However, the Court has not clarified in what cases victims would have Article III standing. *See* Daniel Solove, *In re Zappos: The 9th Circuit Recognizes Data Breach Harm*, PRIVACY & SECURITY BLOG (Apr. 9, 2018), https://teachprivacy.com/in-re-zappos-9th-circuit-recognizes-data-breach-harm/ [https://perma.cc/N6DB-CJ4J].

546. Tsesis, *supra* note 17, at 625 (referring to 18 U.S.C. § 875(c) (2018)).

547. *Id.* (referring to 18 U.S.C. § 2339B(a) (2018)).

548. *Id.*

accounts.[549] Because of the presumption in favor of free speech, injunctions that ban speech by technological measures should only be used in rare cases.[550] In general, the best practices of applying technological measures should be determined by the industry and applied voluntarily.

Scholars have criticized the use of the material support statute for criminal prosecution and argued that it can lead to suppression of protected speech relating to terrorism.[551] The use might suppress news items that are published on social media because it is difficult to discern news about terrorism from terrorist propaganda.[552] These concerns are valid, but criminal prosecution should not be ruled out altogether. It should be limited to cases in which an intermediary refrained from removing severe and clear unprotected speech, or algorithmic recommendations upon actual knowledge.

Arguably, criminal prosecution in such cases can limit speech despite being narrowly tailored. Social media employees receive large volumes of requests and have to make the decision to remove unprotected content. It is difficult to determine whether a specific post is protected or not.[553] Thus, to avoid prosecution, intermediaries might prefer to take down legitimate content and accounts to be on the safe side. However, platforms are also driven by economic incentives, and taking down too much content would result in loss of profits. Thus, the degree to which legitimate expression is chilled should be less extensive than at first glance and reflect a proper balance between free speech and the public's safety.[554]

---

549. *Id.* at 626.

550. *Id.* at 627–28 (referring to rare cases of immediate national emergencies).

551. VanLandingham, *supra* note 209, at 43–44.

552. *Id.* at 39–40.

553. Nick Hopkins, *Revealed: Facebook's internal rulebook on sex, terrorism and violence*, GUARDIAN (May 21, 2017 1:00 PM), https://www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence [https://perma.cc/34JR-BQQ9] ("[T]he volume of work, which means [moderators] often have 'just 10 seconds' to make a decision."); *see also* GILLESPIE, *supra* note 47, at 111–12.

554. Klonick, *supra* note 112, at 1627 ("If a platform creates a site that matches users' expectations, users will spend more time on the site and advertising revenue will increase. Take down too much content and you lose not only the opportunity for interaction, but also the potential trust of users." (footnote omitted)).

c. Public Regulation, Algorithmic Impact Assessment, and Ex Post Enforcement

Terrorist speech fits well in the "data pollution" concept—a term first proposed by Professor Omri Ben-Shahar in related contexts—which compares the harm that speech causes to social institutions with environmental pollution.[555] This analogy was proposed because of similarity to environmental harm, abusive use of data collection, and dissemination of harmful speech that infringes on the public interest.[556] This concept can be applied to terrorist speech because inciting terrorist propaganda leaks into the digital ecosystem, causing fear that disrupts social institutions. The harm of digital data pollution is more systemic, decentralized, and complex relative to traditional harm.[557] Professor Ben-Shahar suggested that devices regulating environmental harm can be used in regulating data pollution.[558] First, there are ex ante regulations that policymakers can utilize for achieving the goal of safety by design. Regulation can limit data collection and the way it is shared and thus limit personalized algorithmic targeting.[559] Yet, unlike environmental harm, data is not toxic per se and it is a challenge to determine in advance which collection of data is beneficial and what constitutes "legitimate" purposes for collection.[560] This solution can reduce the efficiency of recommendation systems altogether and the costs might exceed the benefits.[561] Another

555. *See* Ben-Shahar, *supra* note 481, at 106–07.

556. *See id.*

557. *See* COHEN, *supra* note 165, at 182.

558. *See* Ben-Shahar, *supra* note 481, at 108–10.

559. In the European Union, the General Data Protection Regulation (GDPR) tries to determine the principle of data minimization in advance. *See* Regulation (EU) 2016/679 of the European Parliament and 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), art. 78, 2016 O.J. (L 119) 15. The GDPR protects data of EU citizens, but it applies to non-EU companies that offer goods or services to EU consumers. *See* Michael L. Rustad & Thomas H. Koenig, *Towards a Global Data Privacy Standard*, 71 FLA. L. REV. 365, 365 (2019). Thus, it can affect data protection in the United States and is expected to have a global effect. *Id.* at 365–66. In addition, it is already creating a "Brussels Effect"—a race to the top in data protection standards—as the California Consumer Privacy Act of 2018 (CCPA) demonstrates. *Id.* at 403–05. Similar to the GDPR, this U.S. law also outlines a principle of data minimization in the context of consumer data protection. *Id.* at 404.

560. Ben-Shahar, *supra* note 481, at 133–34.

561. *Id.* at 134.

type of ex ante regulation is directed towards technology and applying best practices of moderation. Trying to determine ahead of time which technology is best to curb a certain problem, however, will cause a chilling effect on innovation that would burden speech. Furthermore, best practices of moderating user content might result in removal of inciting content upon knowledge, but it will not necessarily prevent algorithmic recommendations and targeting of unlawful content.

Other solutions are process-oriented focusing on transparency for algorithmic recommendation systems, even for intermediaries that did not opt into the proposed safe haven regime. One must bear in mind, however, that algorithms are guarded trade secrets; therefore, there are legal difficulties to imposing general transparency obligations.[562] Scholars have proposed a range of mechanisms for promoting algorithmic transparency and accountability.[563] For example, some scholars have argued for promoting nuanced algorithmic transparency, due process, and accountability obligations.[564] Other scholars have argued that the way to achieve transparency is by data protection.[565] Legislation modeled after the European Union's General Data Protection Regulation (GDPR) that protects against automated decisionmaking harm[566] and provides a right for individuals to

562. In the related context, data protection, the EU GDPR requires companies and governments to reveal an algorithm's purpose and the data it uses to make decisions, leading some to infer a right to explanation. *See* Katyal, *supra* note 522, at 106. This right, however, aims to protect a data subject's rights and not the third party. In the United States, such a right does not exist. Accountability in algorithmic programing might be achieved by private industry, rather than public regulation. *Id.* at 107–21. But not all intermediaries are expected to opt in to the safe haven regime or apply a voluntary standard of accountability.

563. *See* Bloch-Wehba, *supra* note 534 (manuscript at 4, 6–7) (reviewing different approaches for algorithmic transparency and arguing that there should be algorithmic transparency in the public sector as part of the law of access to government records and freedom of information because the public as a whole is affected by governmental algorithmic decisions); *see also* Rory Van Loo, *The Missing Regulatory State: Monitoring Businesses in an Age of Surveillance*, 72 VAND. L. REV. 1563, 1563 (2019) (proposing general regulatory monitoring on platforms and of business information and explaining that this monitoring will enhance users' privacy).

564. *See, e.g.*, Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH L. REV. 1, 20 (2014); Zarsky, *supra* note 534, at 1540; Ari Ezra Waldman, *Power, Process, and Automated Decision-Making*, 88 FORDHAM L. REV. 613, 624 (2019).

565. *See, e.g.*, Rustad & Koenig, *supra* note 559, at 453.

566. *See supra* note 559; *see also* Margot E. Kaminski, *The Right to Explanation, Explained*, 34 BERKELEY TECH. L.J. 189, 199 (2019).

receive explanations about the model of algorithms[567] may achieve more transparency and promote procedural justice. However, such legislation focuses on data protection of data subjects and is less suitable to reduce the harm algorithmic recommendations inflict on third parties.

A different way to meet this problem is pre-implementation licensing regime that includes obligations of limited disclosure or a review by the Federal Trade Commission (FTC) or an agency like the Food and Drug Administration (FDA) that can allow protecting against algorithmic incitement to terror.[568] But this broad based legal solution involves profound administrative costs. Furthermore, it might not be fully feasible when learning algorithms are at stake and may hinder innovation.[569] This approach removes the burden from individuals and places it upon the company and the licensors instead. But, in doing so, it creates "a regulatory bottleneck for companies that must move quickly in order to compete."[570] Furthermore, "the focus on documentation and process as ends in themselves elevates a merely symbolic structure to evidence of actual compliance with the law," obscures that algorithmic decisionmaking erodes "substantive values of fairness, equality, and human dignity," and "may thereby discourage both users and policymakers from taking more robust actions."[571]

A superior solution that extends even beyond the design stage is an algorithmic impact assessment that will require intermediaries to ascertain that their algorithms and tools undergo evaluation for safety by independent auditors and technology experts regularly. Algorithmic impact assessment can mitigate the risk for error or failure in the design stage or unexpected reactions of learning algorithms that may result in unprotected recommendations. This idea is not so revolutionary. Recently, legislators proposed an impact assessment for algorithmic dis-

---

567. *See* Andrew D. Selbst & Solon Barocas, *The Intuitive Appeal of Explainable Machines*, 87 FORDHAM L. REV. 1085, 1107 (2018).

568. Andrew Tutt, *An FDA for Algorithms*, 69 ADMIN. L. REV. 83, 115–16 (2017).

569. *See* THIERER ET AL., *supra* note 419, at 18–20.

570. Dennis D. Hirsch, *From Individual Control to Social Protection: New Paradigms for Privacy Law in the Age of Predictive Analytics*, 79 MD. L. REV. (forthcoming 2020) (manuscript at 40).

571. Waldman, *supra* note 564, at 629.

crimination.[572] The proposed bill, the Algorithmic Accountability Act of 2019, requires entities that use, store, or share personal information to conduct impact assessments for automated decision systems and data protection.[573] These impact assessments are meant to monitor for discrimination and give entities a chance to correct discriminatory algorithms in a timely manner.[574] Adding an ex post review to ex ante measures can also be used to promote the public's safety.

However, this solution is not optimal.[575] It still leaves opacity regarding the algorithmic functions and guidelines for implementation measures. Yet, this solution is flexible and might be superior to ex ante full disclosure to the regulator. It can also apply on intermediaries' that chose not to join the safe haven program. Regulators and policymakers are expected to develop clearer guidelines for improving the implementation of this solution.

Ex post public enforcement is another administrative solution.[576] Indeed, liability in tort law might partially compensate victims and their families by applying the loss chances doctrine.[577] But it is more difficult to hold intermediaries responsible for possible harm that might occur in the future.[578] A public enforcement scheme is not constrained by the same remedial standards.[579] A criminal fine, a civil emission fee, or even statutory damages awarded in private class action can lead to deterrence and mitigation of harm. This public regulation can apply to intermediaries that knowingly avoid removing severe unprotected inciting speech.[580] It can also apply to intermediaries that did not opt into the safe haven regime for safe algorithmic recommendations but fail to exercise a duty of care in designing

572. Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. § 3(b) (2019).

573. *Id.*

574. *See id.*; *see also* Margot E. Kaminski & Andrew D. Selbst, *The Legislation That Targets the Racist Impacts of Tech*, N.Y. TIMES (May 7, 2019), https://nyti.ms/2Ybb8MT [https://perma.cc/PZZ7-XV5C]; Waldman, *supra* note 564, at 628–29. The obligation of regular evaluation of algorithmic tools would be enforced by the FTC. *See* Kaminski & Selbst *supra*.

575. *See* Kaminsky & Selbst, *supra* note 574.

576. Ben-Shahar, *supra* note 481, at 47–48.

577. *See supra* Part IV.C.4.c.

578. Solove & Citron, *supra* note 542, at 750.

579. Ben-Shahar, *supra* note 481, at 47–48.

580. *Id.*

recommendation systems and to intermediaries that opted into the safe haven regime but fail to comply with its requirements.

5. Voluntary Prevention and Mitigation

The proposals in this Article outline minimum standards, and they only address unprotected speech or recommendations on such content. Outlining broader mandatory obligations might be unconstitutional. It is also likely to result in extensive collateral censorship and reduce efficiency and innovation.[581] Mandates are not, however, the last word on this topic. In many cases intermediaries can and do mitigate the harm caused by terrorist activities above this minimum legal threshold. They are in a position of responsibility and have an implicit contract with the public to find ways to prevent harm. This social contract does not bind platforms in court, but it is upheld in the court of public opinion.[582] This Part gives a few examples of additional voluntary measures that intermediaries can take to mitigate terrorist activities on their platform.

a. Improving Detection, Enforcement, and Prevention

Intermediaries can and do moderate harmful content proactively and reactively.[583] They use various technologies for efficient removal. They operate moderators and rely on community flagging and technologies.[584] However, the efforts to remove speech ex post might be futile when terrorists, their sympathizers, and the general public share the offensive content and allow it to spread widely on platforms. Voluntary cooperation among social media giants allows them to share unique digital fingerprints that they automatically assign to videos or photos of offensive content that they have removed from their web-

581. *See supra* Part III.B.

582. GILLESPIE, *supra* note 47, at 208; KOSSEFF, *supra* note 223, at 250 ("Although platforms have taken significant steps to meet their obligations under that social contract, they can and should do more.").

583. KOSSEFF, *supra* note 223, at 246 (explaining how YouTube uses machine learning to automatically identify extremist videos and supplements the teams of moderation who manually review videos for violating YouTube's policies); Klonick, *supra* note 112, at 1625–36.

584. GILLESPIE, *supra* note 47, at 74–110 (reviewing the common ways of moderating content).

sites.[585] This allows their peers to identify the same content on their platforms and remove it, thus mitigating the problem of wide dissemination of harmful content. Websites are expected to cooperate with each other if this measure is perceived as "family friendly" and attracts users who are inclined to that environment. In fact, intermediaries already practice this policy in some cases.[586]

To date, the tools of detection have many flaws in interpreting context. Therefore, removal of all replications of text-based expressions should not be used automatically to prevent content from being uploaded to the net. Rather, it should be used for detection, calling attention to the content for human oversight.[587] A limited use of digital fingerprints for detection of repeated harmful content, leaving the decision of removal in the hands of each intermediary, would mitigate the concern of chilling legitimate content. In addition, limiting legal liability for unprotected speech also mitigates the concern for automatic removal of legitimate extremist content by this technology.[588]

Another developing solution is the use of AI to detect terrorist content. Learning algorithms can be useful for efficient pro-

---

585. Rafał Kuchta, *The hash—a computer file's digital fingerprint*, NEWTECH.LAW (Oct. 9, 2017), https://newtech.law/en/the-hash-a-computer-files-digital-fingerprint/ [https://perma.cc/VJZ6-R93G].

586. *See* Julia Fioretti & Lily Cusack, *EU urges internet companies to do more to remove extremist content*, REUTERS (Dec. 6, 2017, 2:14 PM), https://www.reuters.com/article/us-eu-internet-forum/eu-urges-internet-companies-to-do-more-to-remove-extremist-content-idUSKBN1E02Q7 [https://perma.cc/UPH9-3J2V] ("Over the summer, Microsoft (MSFT.O), Facebook, Twitter and YouTube formed a global working group to combine their efforts in removing extremist content from their platforms, and last year formed a database of known 'terrorist' images and videos which now contains more than 40,000 hashes, or digital signatures."); Julia Fioretti, *Web giants to cooperate on removal of extremist content*, REUTERS (Dec. 5, 2016, 6:10 PM), https://www.reuters.com/article/us-internet-extremism-database/web-giants-to-cooperate-on-removal-of-extremist-content-idUSKBN13U2W8 [https://perma.cc/TY99-BH2U]; *see also* Klein & Flinn, *supra* note 17, at 79–81 (referring to the use of this practice in removing child pornography images).

587. This conclusion is reinforced for text-based content—as opposed to images—because the systems today are not very good with handling interpretation and context. *See* GILLESPIE, *supra* note 47, at 98–108.

588. In a recent article, Professor Citron even refers to the use of removal technology as part of a duty of care. Citron, *supra* note 47 ("Internet service providers (ISPs) and social networks with millions of postings a day cannot plausibly respond to complaints of abuse immediately, let alone within a day or two. On the other hand, they may be able to deploy technologies to detect content previously deemed unlawful. The duty of care will evolve as technology improves.").

active detection of such content.[589] These systems are constantly improving,[590] but at this stage they are not good enough in interpreting context.[591] Therefore, any use of automated content analysis tools should be accompanied by human review of the output or conclusions.[592]

Intermediaries and search engines use AI and other technologies to decrease terrorist content visibility, such as livestreaming of terror attacks,[593] or to detect use of hashtags to increase the visibility of harmful content and block them.[594] For example, following the terror attack in New Zealand, Facebook decided to improve its matching technology tools to stop the spread of viral videos of this nature and expand collaboration with the industry to counter terrorism.[595]

Intermediaries can also counter terrorists' posts and mitigate extremism through anti-terror advertising.[596] Jigsaw, one of Google's semi-independent units, has recently taken on the challenge of identifying extremist content of terrorist groups before it erupts into violence.[597] Nevertheless, there are practi-

589. *See, e.g.*, BRUNDAGE ET AL., *supra* note 39, at 60; GILLESPIE, *supra* note 47, at 97; Sheera Frenkel, *Facebook Will Use Artificial Intelligence to Find Extremist Posts*, N.Y. TIMES (June 15, 2017), https://www.nytimes.com/2017/06/15/technology/facebook-artificial-intelligence-extremists-terrorism.html [https://perma.cc/SW3Y-5SVV].

590. *See* Elkin-Koren, *supra* note 366, at 1097.

591. GILLESPIE, *supra* note 47, at 98–108.

592. DUARTE ET AL., *supra* note 440, at 6.

593. For example, algorithms could reduce or obscure the visibility of the live streaming of the terror attack in New Zealand, *see supra* note 18 and accompanying text. Obscuring content is in line with § 230. *See* Force v. Facebook, Inc., 934 F.3d 53, 70 n.24 (2d Cir. 2019) ("We do not mean that Section 230 requires algorithms to treat all types of content the same. To the contrary, Section 230 would plainly allow Facebook's algorithms to, for example, de-promote or block content it deemed objectionable.").

594. *See, e.g.*, Casey Newton, *Instagram will begin blocking hashtags that return anti-vaccination misinformation*, VERGE (May 9, 2019, 12:37 PM), http://www.theverge.com/2019/5/9/18553821/instagram-anti-vax-vaccines-hashtag-blocking-misinformation-hoaxes [https://perma.cc/EC6J-3U7G]. Google's algorithms are subject to regular tinkering from executives and engineers on specific search results, including on topics such as vaccinations and autism. *See* Grind, *supra* note 151.

595. *See* Guy Rosen, *A Further Update on New Zealand Terrorist Attack*, FACEBOOK NEWSROOM (Mar. 20, 2019), https://newsroom.fb.com/news/2019/03/technical-update-on-new-zealand/ [https://perma.cc/2PSL-TLRE].

596. *See* Andy Greenberg, *Google's Clever Plan to Stop Aspiring ISIS Recruits*, WIRED (Sept. 7, 2016, 7:00 AM), https://www.wired.com/2016/09/googles-clever-plan-stop-aspiring-isis-recruits/ [https://perma.cc/AD8Q-QG23].

597. *Id.*

cal challenges in classifying content correctly, resulting in flagging innocent individuals as terrorists. Another challenge is philosophical: by predetermining that someone is a terrorist based on past patterns, AI might infringe on that person's autonomy.[598]

As technology progresses, existing ways of moderation are expected to improve, and become more accurate. This will allow intermediaries to work beyond the minimum standard of unprotected speech and voluntarily mitigate the harm caused by terrorists' propaganda, incitement, and recruitment. Ethical standards and obligations should develop and encourage intermediaries to use data they gain from operating the platform to prevent harm.[599]

Voluntary measures for algorithmic enforcement define the scope of rights without transparency.[600] One possibility to mitigate the problem of algorithmic enforcement is facilitating an out-of-court dispute settlement system to resolve disputes related to the removal or disabling of access to illegal content, as recommended by the European Council.[601] This system will allow users to challenge intermediaries' decisions to take down content. Another strategy is revealing improper speech restrictions by private initiatives that are committed to protect online free speech. Such initiatives would increase the awareness of policymakers, the press, and the public to online free speech violations, and lead to public outcry that would mitigate improper speech restrictions.[602] In addition, intermediaries can voluntarily disclose information on their enforcement practices by transparency reports and allow users to challenge removal decisions.[603]

It should be noted that following public concerns, Facebook is already proposing to create an independent body to make

---

598. GILLESPIE, *supra* note 47, at 109–10; *see also* Gal, *supra* note 154, at 75–76.

599. *See* Maldoff & Tene, *supra* note 479.

600. *See* Balkin, *supra* note 148, at 1167.

601. Commission Recommendation (EU) 2018/334 of 1 March 2018 on measures to effectively tackle illegal content online, 2018 O.J. (L 63) 58 ("Member States are encouraged to facilitate, where appropriate, out-of-court settlements to resolve disputes related to the removal of or disabling of access to illegal content.").

602. Perel & Elkin-Koren, *supra* note 405, at 202–05.

603. Google already publishes transparency reports. *See Google, Transparency Report*, GOOGLE, https://transparencyreport.google.com/?hl=en [https://perma.cc/K2ER-DCU2] (last visited Nov. 5, 2019).

decisions about what kinds of content users would be allowed to post and include an oversight committee.[604] Such a body can help highlight weaknesses in the policy formation of platforms, provide an independent forum for discussing disputed content moderation decisions, and allow public reasoning necessary for users.[605] These measures and others can enhance accountability.

b. Rethinking Legal and Ethical Considerations of Design to Prevent Harmful Outcomes of the Algorithmic Code

Intermediaries can do more to prevent harmful outcomes of algorithmic recommendation.[606] Scholars and even governments have addressed the need for a framework for designers[607] and for an ethical code for code developers[608] in related contexts; the U.K. government has even set up a Center for Data Ethics and Innovation to provide independent advice on the ethical and innovative deployment of data and AI.[609] The industry can develop ethical guidelines as well. Recently, scholarly work has proposed to develop a set of ethical principles within professional organizations like the Association for the Advancement of Artificial Intelligence and the Association of Computing Machinery.[610] The proposed ethical guidelines and principles of algorithmic accountability are applicable to algorithmic recommendations that promote terrorist attacks. They are not aimed at censoring users but rather address intermediaries' algorithmic recommendations that are not always in line with the intermedi-

---

604. On the planned oversight board, its benefits, and limitations, see Evelyn Douek, *Facebook's "Oversight Board:" Move Fast with Stable Infrastructure and Humility*, N.C. J.L. & TECH., Oct. 2019, at 1, 28–49 (2019). This idea can promote the removal of unlawful content above the legal threshold. There are more possibilities that can mitigate harm. *See* Kate Klonick & Thomas E. Kadri, Opinion, *How to Make Facebook's 'Supreme Court' Work*, N.Y. TIMES (Nov. 17, 2018), https://nyti.ms/2Ds8Ba3 [https://perma.cc/NCY4-BWZ7].

605. *See* Douek, *supra* note 604, at 67–68.

606. Mulligan & Bamberger, *supra* note 404, at 701.

607. *See id.* at 742; Levy & Barocas, *supra* note 143.

608. *See* WORLD ECON. FORUM, HOW TO PREVENT DISCRIMINATORY OUTCOMES IN MACHINE LEARNING 21 (2018), http://www3.weforum.org/docs/WEF_40065_White_Paper_How_to_Prevent_Discriminatory_Outcomes_in_Machine_Learning.pdf [http://perma.cc/N9CW-Z5N9] (referring to discriminatory design and proposing "Principles on the Ethical Design and Use of AI and Autonomous Systems").

609. U.K. ONLINE HARMS WHITE PAPER, *supra* note 488, at 26.

610. Katyal, *supra* note 522, at 109.

aries' own policies.[611] To improve their service, the industry in general and every intermediary in particular should identify the values they strive to promote. They should then encourage code developers, engineers, and legal advisors of technology companies to consider the full range of values and public interests implicated by technical design.[612] Considering the values at stake beforehand should enhance accountability in code development, reduce negligent design, and mitigate the harmful consequences of algorithms.

After developing the algorithmic code, ex post impact assessment statement of algorithmic recommendation systems can refine and improve the system's accuracy, fairness, and accountability beyond legal obligation. This assessment is desirable even if legislators fail to adopt obligations of algorithmic accountability and technology companies adopt them voluntary.[613]

CONCLUSION

Social networks and new methods of communication enable users to spread content and find it easily. New digital developments create an ecosystem for terrorists to spread propaganda, recruit and incite others to commit terrorist attacks. Online intermediaries provide platforms and communication tools for the public. They also enhance terrorist activities by targeting personalized recommendations to consume unlawful content and connect with affiliates of FTOs.

This Article addressed the question whether online intermediaries bear responsibility for terror attacks. It argues that the law should react to the change of ecosystem and prosperity of terrorists' content and incitement. Intermediaries use new innovative communication tools, advanced targeting abilities, and new strategies of moderation. They possess great power and influence over online incitement. With greater power should come greater responsibility. Because of the change in the ecosystem of incitement online, policymakers should outline a new balance among norma-

611. *See* WORLD ECON. FORUM, *supra* note 608, at 13–14; Gerrard, *supra* note 38, at 4503–05; Manheim & Kaplan, *supra* note 38, at 147.

612. Mulligan & Bamberger, *supra* note 404, at 701–02.

613. Katyal, *supra* note 522, at 117 (discussing discrimination and learning algorithms and suggesting enlisting engineers to explain their design choices and evaluate their efficacy).

tive considerations for intermediaries' liability. The new balance should account for intermediaries' influences on the flow of information. Policymakers should develop and impose old and new obligations, remedies, and sanctions on intermediaries to mitigate harm caused by terrorism. The Article proposed a minimum standard for removal of unprotected speech and standards of safety by design for mitigating the damage caused by algorithmic recommendations.

The Article also addressed the legal barriers for full compensation. It advocated for the application of the loss chance doctrine in suits filed by terror victims against the worst intermediaries, thus allowing for partial compensation. It further proposed more obligations and sanctions on intermediaries in criminal and civil law. The proposals pose a minimum threshold and do not preclude voluntary measures that intermediaries can take to mitigate harm caused by terrorist activities.