2022

When AI Goes to War: Corporate Accountability for Virtual Mass Disinformation, Algorithmic Atrocities, and Synthetic Propaganda

Jon M. Garon

NORTHERN KENTUCKY LAW REVIEW

NKU SALMON P. CHASE COLLEGE OF LAW
NORTHERN KENTUCKY UNIVERSITY

NORTHERN KENTUCKY LAW REVIEW

Volume 49 General Law Issue Number 2

When AI Goes to War: Corporate Accountability for Virtual Mass Disinformation, Algorithmic Atrocities, and Synthetic Propaganda

*Jon M. Garon**

For the superficial observer, the development of communications media—the network of land, sea and air ways, of postal, telegraphic and telephonic communications, of radio and television—may be a simple matter of economy or a sort of game. In reality, it is a potent phenomenon of nature.[1]

We live "in a world that is being born *instead of a world that* is. *"*[2]

— Piet Smulders

I. Introduction

The internet has the power to destabilize nations and threaten dictators. The Arab Spring uprisings of 2012 offered evidence of the influence social media can put on totalitarian regimes, but the January 6, 2021, U.S. Capitol Insurrection also highlights how social media can be exploited—or even manipulated—into propaganda sufficient to trigger mobs, murders, and revolution aimed at legitimate democratic institutions.

The internet does not, itself, create either the machinery of war or the power of propaganda. Propaganda and the weapons of war are made by legal entities working at the behest of governments or military powers. During World War II, the U.S. War Production Board converted most of U.S. industrial output from domestic products to war production. In the U.S. automotive industry, for example, automobile manufacturing dropped from 3 million vehicles in 1941 to a mere 139 cars in the ensuing three years during the war so that auto manufacturers could produce tanks. The same militarization occurred in Germany.

World War II, however, added the specter of wide-scale crimes against humanity to the industrialization. Three German companies were tried for war crimes following the atrocities of the Holocaust, the most notorious being I.G. Farben, which supplied Zyklon-B to the SS, which was used to exterminate

---

* Professor of Law and Director of the Intellectual Property, Cybersecurity, and Technology Law program, Nova Southeastern University Shepard Broad College of Law. B.A. 1985 University of Minnesota. J.D. 1988 Columbia Law School.

1. Piet Smulders, The design of Teilhard De Chardin 95 (The Newman Press, Westminster 1967).

2. Ian G. Barbour, *Teilhard's Process Metaphysics*, 49 U. Chic. Press 136, 137 (1969) *quoting* Piet Smulders, The Future of Man 261 (1964).

millions of Jews, homosexuals, and others deemed "undesirable" by the Nazi regime.[3] Despite uncontested evidence of the company's production and supply of the poison, neither the company nor its executives were found guilty of war crimes related to the sale of Zyklon-B for its use as the primary agent of genocide. The tribunal instead noted that the rat poison had a commercial use in the concentration camps and that there was insufficient evidence linking the defendants to the knowing use of the poison against the prisoners.

Companies make products and machines that can be used to kill, sometimes on a vast scale. They can also produce rousing rhetoric and misinformation designed to do the same. Propaganda, of course, is certainly not new. But wars waged without professional armies rely heavily on propaganda. "[M]odern wars required propaganda to mobilise hatred against the enemy; to convince the population of the justness of the cause; to enlist the active support and cooperation of neutral countries; and to strengthen the support of allies."[4]

War and genocide, however, are rarely initiated without notice. Instead, the belligerent actors threaten and vilify their enemy to create public tolerance for their atrocities. One way, therefore, to stop war and genocide is to disrupt the propaganda that precedes it.

Unfortunately, in the modern age, the creation of harmful, deceitful propaganda has become easier than ever. The internet has the potential to promote misinformation on an industrial scale with alarming ease. For example, in the past decade, the Rohingya, a Muslim ethnic minority in Myanmar (or Burma), were subjected to a military-led series of attacks designed to promote ethnic cleansing or genocide.[5] Although the ethnic and religious animosity dates back decades, if not centuries, the ability to use Facebook gave the military the tools to promote and carry out its most recent campaign of genocide.[6] Facebook's role in the Rohingya genocide provides merely one example of the global threats that are

---

3. Terese Pencak Schwartz, *The Holocaust: Non-Jewish Victims*, JEWISH VIRTUAL LIBRARY, (last visited Feb. 17, 2022) https://www.jewishvirtuallibrary.org/non-jewish-victims-of-the-holocaust ("Although Jews were the primary victims of the Nazi's evil, many other groups were targeted based on both racial and political grounds ... [including] LGBTQ individuals, the physically and mentally disabled, Roma (gypsies), Poles and other Slavic peoples, Jehovah's Witnesses, and members of political opposition groups.").

4. Propaganda, British Library (last visited Dec. 19, 2021), https://www.bl.uk/world-war-one/themes/propaganda.

5. *See* Betsy Swan, *Exclusive: Facebook Silences Rohingya Reports of Ethnic Cleansing*, DAILY BEAST (Sept. 18, 2017), https://www.thedailybeast.com/exclusive-rohingya-activists-say-facebook-silences-them; Shalailah Medhora, *Number of Rohingyas fleeing Myanmar tops 600,000 since August*, ABC (AU) (Nov. 16, 2017), https://www.abc.net.au/triplej/programs/hack/rohingyas-fleeing-myanmar-tops-600000/9158404

6. *See* Doe v. Meta Platforms, Inc., 3:22-cv-00051, (Cal. Sup. Ct. Jan. 5, 2022) https://www.pacermonitor.com/public/case/43161115/Doe_v_Meta_Platforms, Inc; Betsy Swan, *Exclusive: Facebook Silences Rohingya Reports of Ethnic Cleansing*, DAILY BEAST (Sept. 18, 2017), https://www.thedailybeast.com/exclusive-rohingya-activists-say-facebook-silences-them; Shalailah Medhora, *Number of Rohingyas fleeing Myanmar tops 600,000 since August,* ABC (AU) (Nov. 16, 2017), https://www.abc.net.au/triplej/programs/hack/rohingyas-fleeing-myanmar-tops-600000/9158404.

emerging as social media, metaverses, and digital content displaces first-hand knowledge and journalistic reporting of facts and events. The world has faced this before. Corporate technology is essential for nation-states to wage war, including the disinformation for war. In today's world, internet media technologies are as essential as the tank production of World War II.

In the past, the efforts to fuel racial and ethnic hatred or to justify military nationalism required significant resources and were subject to severe practical limitations. Over the next decade, however, the growth of the metaverse, the expansion of artificial intelligence (AI), and the continued improvements in synthetic media will combine to create a network for communication. Unlike the communications platforms in the age of mass media, the many-to-many media networks can be used to promote stories and perspectives that are difficult to source, and it will be much easier to obfuscate their origins. Without adequate safeguards, the use of avatars, AI, deepfakes, and other tools will make dissemination of misinformation even easier than before. As a result, ostensibly credible disinformation can be exploited by belligerent nations, terrorist networks, and non-state actors to foster territorial conflict and genocide. Worse, these tools may potentially be exploited by individual madmen, unleashing a new scale of public threat.

This article serves to highlight the growing threat of virtual mass disinformation, to identify the need for new regulations at the national level, and to identify areas where international law and treaty must restrict the legality of such actions. Finally, the article proposes that enterprises involved in internet platforms bear a duty of reasonable care to assure that virtual, artificial propaganda does not spread through that enterprise's platform.

II. WHERE WE ARE: SOCIAL RESPONSIBILITY UNDER WEB 2.0

In September 2000, the United Nations held the Millennium Summit, designed to map out the role of the United Nations in the coming century, to commit global resources to reduce global poverty, to help address global health crises, and to improve peace and stability throughout the world.[7] Five years later, the UN held a follow-up meeting at which the UN produced the World Summit Outcome Document,[8] which reiterated many of the core precepts of the UN and the commitments made during the Millennium Summit. It also added, for the first time, the notion that the UN and external sovereign nations had the "responsibility to protect" people from "national authorities manifestly failing to protect their

7. *See* Press Release, *World Leaders Adopt "United Nations Millennium Declaration" At Conclusion Of Extraordinary Three-Day Summit,* UNITED NATIONS (Sept. 8, 2020), https://www.un.org/press/en/2000/20000908.ga9758.doc.html.

8. 2005 World Summit Outcome, G.A. Res. 60/1, ¶¶ 138-139, U.N. Doc. A/Res/60/1 (Sept. 16, 2005) [hereinafter World Summit Outcome Document]. *See generally*, Major Jeremy A. Haugh, *Beyond R2p: A Proposed Test for Legalizing Unilateral Armed Humanitarian Intervention*, 221 MIL. L. REV. 1, 74 (2014).

populations from genocide, war crimes, ethnic cleansing and crimes against humanity."[9]

Responsibility to Protect (R2P or RtoP) was explained in two paragraphs of the World Summit Outcome Document. In § 138, the UN provided that "[e]ach individual State has the responsibility to protect its populations from genocide, war crimes, ethnic cleansing and crimes against humanity. This responsibility entails the prevention of such crimes, including their incitement, through appropriate and necessary means."[10] Section 138 articulates a positive obligation each state has to thwart the incitement to genocide against any of its people, placing this obligation within the larger obligation of each nation to undertake sovereign duties to protect the population within the nation's territory.[11] States' authorities need to ensure that minorities enjoy the fundamental right to equality, both in written legislation and in society at large.[12]

The roles of local government, civic organizations and non-governmental organizations ("NGOs") are important in this respect. Police, prosecutors and judges need to be more aware of what constitutes racial discrimination and racially motivated crimes and, in some cases, changing the composition of police forces to better reflect the multi-ethnic communities they serve may be appropriate. Other recommendations include monitoring hate speech, promoting empowerment through education, and ensuring adequate housing and access to health care.

Section 139 goes much further. It grants the UN the authority and "the responsibility to use appropriate diplomatic, humanitarian and other peaceful means ... to help protect populations from genocide, war crimes, ethnic cleansing and crimes against humanity."[13] In situations where a sovereign nation fails to

---

9. World Summit Outcome Document, *supra* note 8 at 139.

10. *Id.* at §138. The provision, in full, provides as follows:

> 138. Each individual State has the responsibility to protect its populations from genocide, war crimes, ethnic cleansing and crimes against humanity. This responsibility entails the prevention of such crimes, including their incitement, through appropriate and necessary means. We accept that responsibility and will act in accordance with it. The international community should, as appropriate, encourage and help States to exercise this responsibility and support the United Nations in establishing an early warning capability.

11. *See* U.N. Secretary-General, Implementing the Responsibility to Protect: Report of the Secretary-General, ¶ 3, U.N. Doc. A/63/677 (Jan. 12, 2009) [hereinafter Implementing R2P], available at https://www.un.org/ruleoflaw/blog/document/report-of-the-secretary-general-implementing-the-responsibility-to-protect/ ("It should be underscored that the provisions of paragraphs 138 and 139 of the Summit Outcome are firmly anchored in well-established principles of international law. Under conventional and customary international law, States have obligations to prevent and punish genocide, war crimes and crimes against humanity.").

12. *Multi-ethnic States and the Protection of Minority Rights*, UNITED NATIONS WORLD CONFERENCE AGAINST RACISM (last visited Dec. 20, 2021), https://www.un.org/WCAR/e-kit/minority.htm.

13. World Summit Outcome Document, *supra* note 8 at 139. The provision, in full, provides as follows:

> 139. The international community, through the United Nations, also has the responsibility to use appropriate diplomatic, humanitarian and other peaceful

protect the people within its borders, the UN has the power to take actions. The authority is limited to the UN acting as a whole.[14]

The responsibility and authority were recognized by the UN Security Council, which reaffirmed its commitment to the UN taking steps to intervene with nations which failed to protect their populations.[15]

The expansion of the UN's Responsibility to Protect authority suggests a global movement designed to enhance the protections for individuals. The nineteenth and twentieth centuries could be understood as an era of global development through the emergence of the nation-state.[16] Globalization and the internet age have been seen as weakening the dominance of the nation-state.[17] It has even been suggested that "[g]lobalization and the increasing movement of

---

> means, in accordance with Chapters VI and VIII of the Charter, to help protect populations from genocide, war crimes, ethnic cleansing and crimes against humanity. In this context, we are prepared to take collective action, in a timely and decisive manner, through the Security Council, in accordance with the Charter, including Chapter VII, on a case-by-case basis and in cooperation with relevant regional organizations as appropriate, should peaceful means be inadequate and national authorities manifestly fail to protect their populations from genocide, war crimes, ethnic cleansing and crimes against humanity. We stress the need for the General Assembly to continue consideration of the responsibility to protect populations from genocide, war crimes, ethnic cleansing and crimes against humanity and its implications, bearing in mind the principles of the Charter and international law. We also intend to commit ourselves, as necessary and appropriate, to helping States build capacity to protect their populations from genocide, war crimes, ethnic cleansing and crimes against humanity and to assisting those which are under stress before crises and conflicts break out.

14. *See id.*

15. S.C. Res. 1674, ¶ 4, U.N. Doc. S/Res/1674 (Apr. 28, 2006), available at https://www.un.org/ruleoflaw/blog/document/security-council-resolution-1674-2006-on-protection-of-civilians-in-armed-conflict/.

16. *See* Andreas Wimmera and Yuval Feinstein, *The Rise of the Nation-State across the World, 1816 to 2001*, 75 AM. SOC. REV. 764 (2010) DOI: 10.1177/0003122410382639 ("The French and American revolutions of the late-eighteenth century gave birth to the ideal of the modern nation-state—an independent state with a written constitution, ruled in the name of a nation of equal citizens.").

17. RICHARD BATLEY & GEORGE A. LARBI, THE CHANGING ROLE OF GOVERNMENT: THE REFORM OF PUBLIC SERVICES IN DEVELOPING COUNTRIES 2 (2004).

> The latter part of the twentieth century and the early twenty-first century have seen the emergence of a more porous view of the nation state, and changed views of the role of government: it would perform fewer functions on its own and more in partnership with other actors. Associated with this change in public policy about what the state was to do was a consequential change in how it was to act.

See also Austen L. Parrish, *Lands, Liberties, And Legacies: Indigenous Peoples and International Law: Theoretical Approaches to International Indigenous Rights: Changing Territoriality, Fading Sovereignty, and the Development of Indigenous Rights*, 31 AM. INDIAN L. REV. 291, 302 (2007) ("the salience of the sovereign state, strictly defined by its territorial borders, has slowly declined."); *see also* MANUEL CASTELLS, INFORMATION TECHNOLOGY, GLOBALIZATION AND SOCIAL DEVELOPMENT, U.N. RES. INST. SOC. DEV., DISCUSSION PAPER NO. 114, at 4 (Sept. 1999).

people across borders threaten to kill off the nation state once and for all."[18] While this may overstate the threat to sovereign nations, they are becoming less stable.[19]

The challenge to sovereign authority is twofold. First, governments are threatened by the power the individual has gained using technology to engage in extraterritorial activities.[20] Second, governments are threatened by the scale of key multinational informational platforms that have the economic power and informational reach to topple nations, if they so choose.[21] "The UN system has the potential to challenge the hegemony of corporations and the elites who use and manage them. But for the last several decades it has been used to support an elite project of capitalist expansion and neoliberal globalization, enabling the growth of new corporate 'shadow sovereigns.'"[22] The twin challenges presented by

---

18. *Multi-ethnic States and the Protection of Minority Rights*, *supra* note 12.

19. *See* Richard H. Pildes, *Political Fragmentation in Democracies of the West* 1 (N.Y.U. Pub. L. & Legal Theory Working Paper, Paper No. 21-50, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3935012# (discussing political fragmentation as the description of governmental instability and identifying the changes to the communications infrastructure as a leading cause of the political fragmentation).

20. *See* Riel Miller, Wolfgang Michalski & Barrie Stevens, *The Promises and Perils of 21st Century Technology: An Overview of the Issues, in* 21ST CENTURY TECHNOLOGIES, PROMISES AND PERILS OF A DYNAMIC FUTURE 7, 11 (Organisation For Economic Co-Operation and Development ed., 1998) ("Anyone with a computer and Internet access will be in a position to become a merchant and reach out to customers across the globe, and any consumer will be able to shop the world for goods and services."); *see also* Daren Fonda, *Why Cryptocurrencies Are a Threat to Central Banks*, BARRON'S (May 3, 2021), https://www.barrons.com/articles/cryptocurrency-is-threatening-the-role-of-central-banks-why-governments-must-go-crypto-51619814196 (use of cryptocurrencies allows individuals to avoid money transfer restrictions); *see also* Andrea Carson, *How Investigative Journalists are Using Social Media to Uncover the Truth*, THE CONVERSATION (Oct. 17, 2016, 3:20 PM), https://theconversation.com/how-investigative-journalists-are-using-social-media-to-uncover-the-truth-66393 (investigative reporting).

21. *See* Jackie Smith, *Challenging Corporate Power: Human Rights Globalization from Above and Below*, 64 SOC'Y FOR INT'L DEV. 63, 64 (2021), https://link.springer.com/content/pdf/10.1057/s41301-021-00292-2.pdf ("Today's biggest challenges—most notably inequality, environmental collapse, and growing violence—can all be linked at least in part to the problem of corporate power."); *see also* Walter Frick, *The Conundrum of Corporate Power*, HARV. BUS. REV. (May–June 2018), https://hbr.org/2018/05/the-conundrum-of-corporate-power ("A significant body of research suggests that the biggest organizations in most industries account for a larger percentage of revenues and profits in their markets than they did a decade or two ago and that their power has grown."); *see also* BRINK LINDSEY & STEVEN M. TELES, THE CAPTURED ECONOMY 5 (Oxford Univ. Press 2017) (discussing regulatory capture as cause for economic inequality); *see also* Corina Rodríguez Enríquez, *Corporate Power: A Risky Threat Looming over the Fulfilment of Women's Human Rights*, SOCIAL WATCH (last visited Dec. 17, 2021), https://www.socialwatch.org/node/17687 ("there is globally 'a growing reliance on corporate-led solutions to global problems.' But in the context of financialized globalization and the promotion and dominance of self-regulation, it is fair to ask whether the private sector contributes more to the problems than to their solutions.") (quoting BARBARA ADAMS & JENS MARTENS, FIT FOR WHOSE PURPOSE? PRIVATE FUNDING AND CORPORATE INFLUENCE IN THE UNITED NATIONS 5 (Wolfgang Obenland, Karolin Seitz, Eleonora Hoffman, Johannes Peter, Katherine Marshall, Lisa Monschau, Sabá Loftus & Karen Judd eds., Glob. Pol'y F. 2015), www.globalpolicy.org/images/pdfs/images/pdfs/Fit_for_whose_purpose_online.pdf).

22. Smith, *supra* note 21.

increasingly powerful corporations and increasingly powerful communications networks come together in the sphere of public information.

As social media has expanded, these two destabilizing forces combine.[23] In recent years, the phenomenon of the "pop-up" digital political party has emerged to empower anyone with a popular message to quickly form a political party and obtain a platform.

> These pop-up digital parties use technology to promise a new vision of grassroots democracy. They profess to use the digital revolution to offer a form of organizing politics, and political parties, that is more participatory — 'more democratic, more open to ordinary people, more immediate and more direct, more authentic and transparent.'[24]

In many cases, however, the claims of transparency and democracy are often fig leaves for manipulative, artificial, or autocratic campaigns that use the lack of accountability to propose extremist political agendas.[25] The internet has always had a problem with astroturfing, the ability to project a manufactured illusion of grassroots support.

A December 2021 *New York Times* report highlights the extent to which China has undertaken such efforts.[26] "China's government has unleashed a global online campaign to burnish its image and undercut accusations of human rights abuses."[27] The report shows how Chinese leaders "[f]lood global social media with fake accounts used to advance an authoritarian agenda. Make them look real and grow their numbers of followers. Seek out online critics of the state — and find out who they are and where they live."[28] Even though social media giants Facebook and Twitter are officially blocked inside China, they are still the subject of significant government manipulation to stifle dissent and promote the country's ideological agenda. Also in December 2021, Facebook blocked six surveillance companies along with a "mysterious Chinese law enforcement supplier" because the seven surveillance firms were linked to illegal surveillance in over 100 countries involving more than targeted 50,000 targeted accounts.[29] In addition, emerging

---

23. *See* Pildes, *supra* note 19, at 37.

24. *Id.* at 40-41 (quoting PAOLO GERBAUDO, THE DIGITAL PARTY 4 (Jodi Dean, Joss Hands & Tim Jordan eds., Pluto Press 2019)).

25. *Id.* at 43 ("But as is widely known by now, this image of bottom-up, organic, participatory democracy is at best an illusion, at worst, a cynical manipulation by the movement's leaders.").

26. Muyi Xiao, Paul Mozur & Gray Beltran, *Buying Influence: How China Manipulates Facebook and Twitter*, N.Y. TIMES (Dec. 20, 2021), https://www.nytimes.com/interactive/2021/12/20/technology/china-facebook-twitter-influence-manipulation.html.

27. *Id.*

28. *Id.*

29. Thomas Brewster, *Facebook Warns 50,000 Users Were Targeted By Spy-For-Hire Companies*, FORBES (Dec. 16, 2021, 3:00 PM), https://www.forbes.com/sites/thomasbrewster/2021/12/16/facebook-warning-50000-users-they-were-targeted-by-surveillance-for-hire-companies/?sh=691fb659427b.

technologies and Web3 enterprise models using decentralized autonomous organizations (DAOs) may further the risk that ostensibly democratic and transparent activities create the illusion of engagement when they, in fact, are tightly controlled by centralized authority.[30]

The internet and its related technologies pose a unique set of challenges for governance at the national level and at the level of international collaboration. The governments of the world are challenged by the resources and freedoms afforded to the public through the internet. Governments may lose tax revenue. Governments may struggle to stop the public from behaviors perceived as harmful, such as gambling, sales of restricted drugs, or criminal conspiracies. In theocratic states, governments may lose control over religious doctrine and blasphemous content. And in antidemocratic regimes, governments may struggle to control the media or the official version of facts and events.

> Social media allows us to connect across borders, to communicate more easily than at any other time in human history, and even to expose human rights abuses in faraway places. But in this new digital era, Facebook, Instagram, YouTube, Twitter and other platforms are not just places for information sharing and social networking, they are also places where vilification, targeting and incitement take place. Hate speech is not only proliferating in the dark corners of the internet, it is increasingly common on all major social media platforms.[31]

One of the key obligations of the state is to protect the public from the threat of harm by others. Criminal laws, tort laws, and even contract laws are all designed to protect the interests of one person from being usurped by another. The UN took another small step to address this problem in 2011 by adopting the *UN Guiding Principles on Business and Human Rights*,[32] which are "a set of guidelines for States and companies to prevent, address and remedy human rights abuses committed in business operations."[33]

The modern social media system is entirely operated through corporate agents who serve—instead of government agencies—to support the network of

---

30. *See, e.g.*, Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934: The DAO, Release No. 81207, 117 SEC Docket 745 (July 25, 2017), [hereinafter SEC DAO Report] ("The voting rights afforded DAO Token holders did not provide them with meaningful control over the enterprise, because (1) DAO Token holders' ability to vote for contracts was a largely perfunctory one; and (2) DAO Token holders were widely dispersed and limited in their ability to communicate with one another.").

31. *Simon Adams, Glob. Ctr. for the Resp. to Protect, Hate Speech and Social Media: Preventing Atrocities and Protecting Human Rights Online (Feb. 16, 2020).*

32. *See* John Ruggie (Special Representative of the Secretary-General), *Guiding Principles on Business and Human Rights: Implementing the United Nations "Protect, Respect and Remedy" Framework*, U.N. Doc. A/HRC/17/31 (March 21, 2011) (endorsed by the U.N. Human Rights Council resolution 17/4, June 16, 2011).

33. *UN Guiding Principles*, BUS. & HUM. RTS. RES. CTR., https://www.business-humanrights.org/en/big-issues/un-guiding-principles-on-business-human-rights/ (last visited Jan. 22, 2022).

communication and engagement. In each country where the social platforms operate, the government can pass laws to criminalize certain behaviors where one individual or group of individuals harms another or threatens to do so. Those same governments can also create civil or tort liability for such harms. The corporations cannot enact such laws. But the corporations can use their contractual power to require that the users of their systems adhere to the domestic laws.

There has been a great deal written about the failure of large platforms to forestall hate speech and the importance of the largest social media platforms to do much more to improve the online experience for all users.[34] While the debate regarding the appropriate balance among fundamental interests of free speech, privacy rights, and social harms may be intractable, the recent role of social media in the genocide of the Rohingya minority in Myanmar provides an example of what goes wrong when both the state and the corporate actors fail to meet these obligations.

Facebook's social media plays a predominant role in Myanmar. "Experts describe Facebook's role in the country as the *de facto* internet."[35] The reports from Myanmar in 2017 described "waves of Facebook-based misinformation and propaganda aimed at fueling anti-Rohingya fervor, including fabricated reports that families were setting fire to their own homes in an attempt to generate sympathy."[36] In the ethnic conflict, the Myanmar military, which has since declared martial law, characterized the Rohingya as illegal settlers without the right to reside in the country despite evidence of settlement since the twelfth century.[37] Through the military action and the disinformation campaigns, over 600,000 refugees have been forced to flee the county.[38]

---

34. *See* Agnieszka McPeak, *Platform Immunity Redefined*, 62 WM. & MARY L. REV. 1557, 1613 (2021); *see also* Julie E. Cohen, *Law for the Platform Economy*, 51 U.C. DAVIS L. REV. 133, 144-45 (2017); *see also* Danielle Keats Citron & Benjamin Wittes, *The Internet Will Not Break: Denying Bad Samaritans § 230 Immunity*, 86 FORDHAM L. REV. 401, 416 (2017); *see also* Danielle Keats Citron, *Extremist Speech, Compelled Conformity, and Censorship Creep*, 93 NOTRE DAME L. REV. 1035 (2018); *see also* Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 HARV. L. REV. 1598 (2018); *see also* Alan Z. Rozenshtein, *Surveillance Intermediaries*, 70 STAN. L. REV. 99 (2018); *see also* Mark A. Lemley, *Rationalizing Internet Safe Harbors*, 6 J. ON TELECOMM. & HIGH TECH. L. 101, 102 (2007); JAMES BOYLE, SHAMANS, SOFTWARE, AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY (Harvard Univ. Press 1996).

35. Ingrid Burrington, *Could Facebook Be Tried for Human-Rights Abuses?*, THE ATLANTIC (Dec. 20, 2017), https://www.theatlantic.com/technology/archive/2017/12/could-facebook-be-tried-for-war-crimes/548639/.

36. Mathew Ingram, *In Some Countries, Fake News on Facebook is a Matter of Life and Death*, COLUM. JOURNALISM REV. (Nov. 21, 2017), https://www.cjr.org/analysis/facebook-rohingya-myanmar-fake-news.php.

37. *See* Areina Ismail, *7 Facts About the Rohingya Genocide*, THE BORGEN PROJECT (June 29, 2018), https://borgenproject.org/seven-facts-about-the-rohingya-genocide/ ("[E]vidence of a 1799 document shows that the Rohingya have resided in Myanmar since the 18th century and possibly earlier, considering the earliest records of Muslims in Myanmar are from the 12th century. Today, there are 1.1 million Rohingya living in Buddhist Myanmar.").

38. *See id.*

Facebook has been accused of fueling the genocide in two ways: by helping the military and Buddhist majority population spread false propaganda against the Rohingya and by systematically removing factual content posted by the victims of the ethnic cleansing.[39] In December 2021, a class action lawsuit was filed against Facebook in California state court alleging negligence and strict product liability.[40] The anonymous lead plaintiff in the lawsuit is a Rohingya refugee presently living in Illinois. The lawsuit faces numerous questions regarding First Amendment protections and statutory limitations. It also highlights the lack of international law regarding corporate culpability for institutions that foster genocide or war crimes by others, a failure that was highlighted during the Nuremberg trials following World War II.

Facebook is certainly not alone in receiving criticism for its failure to stop its corporate resources from being used for genocide and human rights abuses.[41] Oil, mining, security, air services, computer systems, telecommunications, and heavy earth-moving equipment have all been identified as part of the infrastructure that are used by governments, and which can become the agents of insurgency groups or belligerent nations for acts of war, genocide, or human rights abuses.[42]

III. NEW TOOLS FOR CYBERWAR'S DISINFORMATION AND PROPAGANDA

The current genocide in Myanmar and other atrocities being committed against minority populations around the world sometimes take advantage of social media to fuel ethnic tensions, promote nationalism, and fuel calls for territorial expansion. Certainly much more could be done to reduce the current risk. But if the current level of human rights abuses is already unacceptably high, then the United Nations and its member states must consider how much worse developing technologies could make the situation. A short exploration of technologies in the pipeline should serve as a wake-up call for the need for much greater collaboration and legal constraints on the means of waging war and ethnic cleansing.

First, corporate technology is essential for nation-states to wage war. As previously mentioned, during World War II, the U.S. War Production Board converted most of U.S. industrial output from domestic products to war

---

39. *See* Swan, *supra* note 5; *see also* Medhora, *supra* note 5; *see also* Burrington, *supra* note 35; *see also*, Ingram, *supra* note 36.

40. Notice by Defendant Meta Platforms, Inc. of Removal of Class Action at 6, Doe v. Meta Platforms, Inc. (f/k/a Facebook, Inc.), No. 3:22-cv-00051 (N.D. Cal. Jan. 5, 2022)

41. *See* Danielle Olson, *Corporate Complicity in Human Rights Violations Under International Criminal Law*, INT'L HUM. RTS. L. J., 2015, at 1, 1 ("[C]orporations . . . have been recognized as having the potential to impact a wide array of human rights in a variety of industry sectors including: extractive industries, pharmaceutical and chemicals, defense, utility and infrastructure, food and beverages, and Information Technology hardware and telecommunications.").

42. INT'L COMM'N OF JURISTS, CORPORATE COMPLICITY & LEGAL ACCOUNTABILITY 22 (Vol. 2, 2008), http://www.icj.org/wp-content/uploads/2012/06/Vol.2-Corporate-legal-accountability-thematic-report-2008.pdf; *see also* JENNIFER ZERK, CORPORATE LIABILITY FOR GROSS HUMAN RIGHTS ABUSES, https://www.ohchr.org/documents/issues/business/domesticlawremedies/studydomesticelawremedies.pdf (last visited Jan. 28, 2022).

production. In the U.S. automotive industry, for example, automobile manufacturing dropped from 3 million vehicles in 1941 to a mere 139 cars in the ensuing three years during the war.[43] The same militarization occurred in Germany.[44] World War II, however, added the specter of wide-scale crimes against humanity to the industrialization.

> [Corporate executives,] whose companies had collectively smelted steel for tanks and purified aluminum for gunbarrels, formulated the synthetic rubber and gasoline necessary for tires and engines, built airplanes and U-boats and V-2 rocket circuit boards, and manufactured nerve gas and Zyklon B. They had seized Jewish property and swallowed up businesses sold off for pennies by those fleeing Nazi persecution. They had contracted with the German government to exploit the labor of concentration camp internees and sited factories with the specific goal of better leveraging this free and disposable workforce. They had planned, profited from, and above all else made possible the Nazi war machine and its genocides.[45]

Only three German companies were ultimately tried for war crimes following the atrocities of the Holocaust,[46] the most notorious being I.G. Farben, which supplied Zyklon B to the SS used in the concentration camps to improve the efficiency for the murder of those incarcerated.[47] Despite uncontested evidence of the company's production and supply of the poison, neither the company nor its executives were found guilty of war crimes related to the sale of Zyklon B for its use as the primary agent of genocide.[48] Although the atrocities of World War II are

---

43. David Vergun, *During WWII, Industries Transitioned From Peacetime to Wartime Production*, U.S. DEPT. OF DEFENSE (Mar. 27, 2020), https://www.defense.gov/News/Feature-Stories/story/Article/2128446/during-wwii-industries-transitioned-from-peacetime-to-wartime-production/.

44. Erica X Eisen, *The Other Nuremberg Trials, Seventy-Five Years On*, BOSTON REV. (Mar. 22, 2021), https://bostonreview.net/articles/erica-x-eisen-nuremberg/ ("Links between the world of big business and the Nazis were extensive: over 50 percent of companies listed on Berlin's stock exchange in 1932 had significant ties to the Nazi Party, and they experienced a boom in stock value after Hitler seized power the following year.").

45. *Id.*

46. *Id.* ("By 1947 the U.S. legal team in Germany had narrowed its focus to the actions of only three companies: IG Farben, Krupp, and Flick KG.").

47. THE UNITED NATIONS WAR CRIMES COMM'N, LAW REPORTS OF TRIALS OF WAR CRIMINALS (Vol. 10, 1949) (discussing the I.G. Farben and Krupp trials), https://tile.loc.gov/storage-services/service/ll/llmlp/Law-Reports_Vol-10/Law-Reports_Vol-10.pdf; *see also* Mark E. Spicka, *The Devil's Chemists on Trial: The American Prosecution of I. G. Farben at Nuremberg*, 61 THE HISTORIAN 865 (1999).

48. The Commission stated:

> The proof was convincing that large quantities of Zyklon-B had been supplied by the Degesch to the S.S. and that it was actually used in the mass extermination of inmates of concentration camps, including Auschwitz. But neither the volume of production nor the fact that large quantities were destined to concentration camps was in itself sufficient to impute criminal responsibility, as it was

more than 75 years ago, the importance of the corporate sponsors continues to provide a chilling reminder about the power of corporations to use their political will to wage war.

In the twenty-first century, concerns about industrial complicity and misuse of technology focus on other arenas. Specifically, the expanded use of artificial intelligence, media falsification software, and virtual worlds each have the potential for tremendous misuse. Singly and together, these three technologies can vastly increase the scale of human rights abuses and war crimes. There is a significant threat that each of these technologies can be used to wage effective disinformation campaigns and lead to risks of military conflict, radicalization, and human rights abuses. As discussed below, the even greater concern is how these three technologies might be used together to automate war and ethnic cleansing.

A. Artificial Intelligence and Algorithmic Decision Systems

The expansion and reliance on artificial intelligence technologies are among the greatest advances in the twenty-first century and greatest threats for misuse.[49] "In general terms, AI refers to a broad field of science encompassing not only computer science but also psychology, philosophy, linguistics and other areas. AI is concerned with getting computers to do tasks that would normally require human intelligence."[50] More broadly, AI embodies fields of computer science, computer processes that include machine learning, deep learning, big data, and similar labels for complex decision-making. Although the origins of AI focus on machines that replicate human-like choice,[51] the critical nature of AI for purposes of risk

---

> established by the evidence that there existed a great demand for insecticides wherever large numbers of displaced persons, brought in from widely scattered regions, were confined in congested quarters lacking adequate sanitary facilities.

THE UNITED NATIONS WAR CRIMES COMM'N , *supra* note 47 at 24; *see also* Kaushik Das Gupta, *Nazi's Industrial Jackal*, DOWNTOEARTH (Oct. 15, 2014), https://www.downtoearth.org.in/coverage/nazis--industrial-jackal-46676 ("IG Farben was the single largest donor to the election campaign of Adolph Hitler in the late 1920s. . . . Farben produced chemical weapons for the German military and looted chemical industries of the countries Germany occupied during the war."); *see also* Edmund L. Andrews, *THE BUSINESS WORLD; I.G. Farben: A Lingering Relic of the Nazi Years*, N.Y. TIMES (May 2, 1999), https://www.nytimes.com/1999/05/02/business/the-business-world-ig-farben-a-lingering-relic-of-the-nazi-years.html (discussing the continued operations of the company despite the efforts to break it up and liquidate it for its Holocaust atrocities).

49. This is not the threat of the "singularity," the term given for the moment when autonomous machines become self-aware and decide to supplant humanity as the dominant intelligence on the planet.

50. Stefan van Duin & Naser Bakhshi, *Part 1: Artificial Intelligence Defined*, DELOITTE. (Mar. 2017), https://www2.deloitte.com/se/sv/pages/technology/articles/part1-artificial-intelligence-defined.html.

51. *See* A.M. Turing, *Computing Machinery and Intelligence*, 49 Mind 433 (1950), https://www.csee.umbc.edu/courses/471/papers/turing.pdf (introducing the "imitation game" that posed a test designed to determine whether a computer could fool an interrogator into believing the computer was human, widely considered the first implementation of conceptual general artificial

management is that the automated decision-making permitted by the technology allows for decision-making that is largely unsupervised by humans with supervisory authority.[52] AI is often used as a general label for discrete technologies, though those technologies can be used in combination.

> Artificial Intelligence (AI) is the science and engineering of making intelligent machines, especially intelligent computer programs. Machine learning is a circle within AI, one that provides systems the ability to automatically learn and improve from experience without being explicitly programmed through access to data. One way to do this is by using Deep Learning, which is a circle inside Machine Learning. Deep Learning uses algorithms inspired by the structure and function of the brain, called artificial neural networks, to make the programs learn through data analysis.[53]

"Known as 'ADS' (algorithmic decision systems), ADS often rely on the analysis of large amounts of personal data to infer correlations or, more generally, to derive information deemed useful to make decisions."[54] In this Article, AI and ADS are used interchangeably to reflect their common usage across the literature.

These AI systems are used to provide facial recognition software; improve photo-editing technologies; permit self-parking cars (and eventually self-driving cars); help fix grammar and punctuation; animate images; stream videos; assign credit ratings; predict health outcomes; identify cancerous growths in x-rays;

---

intelligence); *see also* Rockwell Anyoha, *The History of Artificial Intelligence*, HARV. SCIENCE IN THE NEWS (Aug. 28, 2017), https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/; *see also* Beatriz Guillén Torres, *The True Father of Artificial Intelligence*, BBVA OPEN MIND (Sept. 4, 2016), https://www.bbvaopenmind.com/en/technology/artificial-intelligence/the-true-father-of-artificial-intelligence/ (In 1956, John McCarthy organized a conference at Dartmouth where "he first coined the term 'artificial intelligence,' defined as the science and engineering of making intelligent machines.").

52. *See* CLAUDE CASTELLUCCIA & DANIEL LE MÉTAYER, UNDERSTANDING ALGORITHMIC DECISION-MAKING: OPPORTUNITIES AND CHALLENGES 1 (Mar. 2019) https://www.europarl.europa.eu/RegData/etudes/STUD/2019/624261/EPRS_STU(2019)624261_EN.pdf ("Human intervention in the decision-making may vary, and may even be completely out of the loop in entirely automated systems. In many situations, the impact of the decision on people can be significant, such as: access to credit, employment, medical treatment, judicial sentences, etc."); *see also* Andrew Lohn, Andrew Parasiliti & William Welser IV, *Should We Fear an AI Arms Race? Five Reasons the Benefits of Defense-Related Artificial Intelligence Research Outweigh the Risks—for Now*, DEFENSE ONE (Feb. 8, 2016), http://www.defenseone.com/ideas/2016/02/should-we-fear-ai-arms-race/125670/ ("Deputy Defense Secretary Robert Work, has said 'We believe strongly that humans should be the only ones to decide when to use lethal force. But when you're under attack, especially at machine speeds, we want to have a machine that can protect us.'").

53. Vishal Thakur, *What Is The Difference Between Deep Learning And Artificial Intelligence?*, SCIENCE ABC (Jan. 19, 2022), https://www.scienceabc.com/innovation/what-is-the-difference-between-deep-learning-and-artificial-intelligence.html.

54. CASTELLUCCIA & LE MÉTAYER, *supra* note 52, at 1.

predict the flight-path of missiles; and much more.[55] Some analysts suggest we are already past the tipping point where the proliferation of AI-based decision-making has supplanted the ability of human institutions to control it.[56]

Examples abound. Anti-ballistic missile systems could be designed to self-identify the source of missile launch sites and automatically respond with missile attacks on those sites.[57] Large-scale weapons systems can be trained to identify ships, planes, tanks, and other military targets in order to target and destroy those objects autonomously. Small-scale micro-drones have the potential to be used as anti-personnel devices, used to degrade communications, fuel, and support networks, or to identify other specific targets using facial recognition, biometric information, or other identification triggers. Although these possibilities are largely precluded in the United States by current military policy, for other countries and for non-country actors, these systems pose a global threat.[58] In many cases, the technology at the heart of these systems aimed at the U.S. and its allies comes from Silicon Valley.

Artificial intelligence is nearing its capacity to provide technological systems the ability to make lightning-fast decisions on critical life-and-death situations, to propagate aggressive cyberattacks against state and non-state actors, and to create falsified information through deepfakes and media manipulation that can lead to civil unrest or the advent of warfare. Other systems have become sufficiently advanced that states can now deploy fully autonomous machines with both defensive and offensive capability.

The Fog of War reflects the truism that "[w]ar is the realm of uncertainty; three quarters of the factors on which action is based are wrapped in a fog of greater or lesser uncertainty."[59] Battles are defined by imperfect information, difficulty in communications, and enemy deception. Into this mix, AI technologies can be both

---

55. *See, e.g.*, *What is Artificial Intelligence and How is it Used?*, EUROPEAN PARLIAMENT NEWS (Mar. 29, 2021), https://www.europarl.europa.eu/news/en/headlines/society/20200827STO85804/what-is-artificial-intelligence-and-how-is-it-used (listing shopping, advertising, web search, digital personal assistants, machine translation, smart homes, cities and infrastructure, cars, cybersecurity, health care, fighting disinformation, health, transportation, manufacturing, food and farming, and public administration).

56. *See* Nisha Talagala, *Don't Worry About The AI Singularity: The Tipping Point Is Already Here*, FORBES (June 21, 2021, 4:16 PM), https://www.forbes.com/sites/nishatalagala/2021/06/21/dont-worry-about-the-ai-singularity-the-tipping-point-is-already-here/?sh=5d009bfa1cd4 ("AI has reached a Tipping Point. . . . where a technology grows and permeates our lives very rapidly, building upon itself.").

57. *See* Gerrit De Vynck, *The U.S. Says Humans will Always be in Control of AI Weapons. But the Age of Autonomous War is Already Here*, WASH. POST (July 7, 2021, 10:00 AM), https://www.washingtonpost.com/technology/2021/07/07/ai-weapons-us-military/ ("According to a U.N. group of weapons and legal experts appointed to document the conflict, drones that can operate without human control 'hunted down' [Libyan strongman Khalifa] Hifter's soldiers as they fled [while retreating from U.N. forces from Turkey].").

58. *See* John Bowden, *Top US General Warns Against Rogue Killer Robots*, THE HILL (July 18, 2017, 11:05 PM), https://thehill.com/policy/defense/342659-top-us-general-warns-against-rogue-killer-robots.

59. Quote attributed to Military Philosopher Carl Von Clausewitz (1780-1831).

a blessing and a curse. "On the rewards side of AI weapons, increased precision is expected to reduce collateral damage, increased speed may stop some attacks before they happen, and autonomy should remove soldiers from the battlefield."[60] A drone strike in the concluding days of the U.S. occupation of Afghanistan, for example, has been attributed to "confirmation bias" by the human operators of the drone and the military responsible for the approval of the attack.[61] Potentially, an effective AI system would not have shared the bias as the human operators, avoiding the civilian loss of life.

Algorithmic decision systems can be vetted to reduce false positives, the computer equivalent of confirmation bias, and as a result, help reduce civilian casualties. In classic military environments between sets of warring troops, the use of object recognition software (of which faces are just one type of object) can be configured to identify military uniforms in order to reduce civilian casualties.[62] Another example is "[l]everaging AI to detect risk to civilian infrastructure in conflict areas, and take steps to reduce that risk through more precise use of force and identifying alternatives. This will avoid longer-term negative effects—like loss of power, water, and food supply—to local populations."[63]

At the same time, however, when AI goes to war, there are a multitude of new areas of uncertainty.[64] In the last three decades, urban terrorism has been fueled by suicide bombers and terrorists planting improvised explosive devices (IEDs). Drones are now widely available, and small forces have already begun to weaponize them for military use.[65] Other projects have trained these drones to be

---

60. Lohn, Parasiliti & Welser IV, *supra* note 52.

61. Eric Schmitt reported:

> The higher-level inquiry into the Kabul strike by the Air Force's inspector general, Lt. Gen. Sami D. Said, blamed a series of erroneous assumptions, made over the course of eight hours as U.S. officials tracked a white Toyota Corolla through the Afghan capital, for causing what he called 'confirmation bias,' leading to the attack.

Eric Schmitt, *No U.S. Troops Will be Punished for Deadly Kabul Strike, Pentagon Chief Decides*, N.Y. TIMES (Dec. 13, 2021), https://www.nytimes.com/2021/12/13/us/politics/afghanistan-drone-strike.html.

62. *See* Larry Lewis, *AI-4-Good in War*, JUST SECURITY (May 15, 2018), https://www.justsecurity.org/56282/ai-4-good-war/ ("Thinking creatively, there are ways that AI can improve decision-making and better protect civilians in armed conflict, because of its ability to process large sets of data and rapidly integrate disparate information sources for humanitarian benefits.")

63. *Id.*

64. *See* Kai-Fu Lee, *The Third Revolution in Warfare*, THE ATLANTIC (Sept. 11, 2021), https://www.theatlantic.com/technology/archive/2021/09/i-weapons-are-third-revolution-warfare/620013/ ("The evolution from land mines to guided missiles was just a prelude to true AI-enabled autonomy—the full engagement of killing: searching for, deciding to engage, and obliterating another human life, completely without human involvement.").

65. Nolan Peterson, *Small Weaponized Drones Are the New IEDs of the Middle East, Top US General Warns*, COFFEE OR DIE MAG. (Feb. 10, 2021), https://coffeeordie.com/weaponized-drones/ ("On the entrenched battlefield of eastern Ukraine, both the Ukrainians and their Russian enemies have jury-rigged and weaponized off-the-shelf small drones to create cheap strike platforms capable

flown by AI systems.[66] The combination of these technologies can already create a lethal fighting force that increases the death toll and reduces the cost to wage war. Add facial recognition tools to target individual targets, and aggressors have a new front for terrorism and regime destabilization.[67]

In 2015, at the International Joint Conference on Artificial Intelligence, participants published an open letter urging that governments and industries restrict the use of AI in autonomous weapons.[68] The letter was signed by notable scientists such as Stephen Hawking and entrepreneurs, including Tesla's Elon Musk.[69] The open letter laid out the risk in stark terms:

> If any major military power pushes ahead with AI weapon development, a global arms race is virtually inevitable, and the endpoint of this technological trajectory is obvious: autonomous weapons will become the Kalashnikovs of tomorrow. Unlike nuclear weapons, they require no costly or hard-to-obtain raw materials, so they will become ubiquitous and cheap for all significant military powers to mass-produce. It will only be a matter of time until they appear on the black market and in the hands of terrorists, dictators wishing to better control their populace, warlords wishing to perpetrate ethnic cleansing, etc. Autonomous weapons are ideal for tasks such as assassinations, destabilizing nations, subduing populations and selectively killing a particular ethnic group. We therefore believe that a military AI arms race would not be beneficial for humanity.[70]

Lethal autonomous weapons systems are now being deployed by Turkish forces and may also be in use elsewhere.[71]

---

of dropping grenades and homemade antipersonnel explosives on their enemies."); *see also* Lee, *supra* note 64.

66. Evan Ackerman, *AI-Powered Drone Learns Extreme Acrobatics*, IEEE SPECTRUM (Oct. 7, 2020), https://spectrum.ieee.org/ai-powered-drone-extreme-acrobatics.

67. *See* Lee, *supra* note 64 (An AI drone "nearly killed the president of Venezuela in 2018, and could be built today by an experienced hobbyist for less than $1,000. All of the parts are available for purchase online, and all open-source technologies are available for download. This is an unintended consequence of AI and robotics becoming more accessible and inexpensive. Imagine, a $1,000 political assassin!").

68. *See Autonomous Weapons: an Open Letter from AI & Robotics Researchers*, FUTURE OF LIFE INST. (July 28, 2015) https://futureoflife.org/2016/02/09/open-letter-autonomous-weapons-ai-robotics/ [hereinafter *Open Letter*].

69. *See* Bowden, *supra* note 58.

70. *Open Letter*, *supra* note 68.

71. *See* Joe Hernandez, *A Military Drone With A Mind Of Its Own Was Used In Combat, U.N. Says*, NAT. PUB. RADIO (June 1, 2021), https://www.npr.org/2021/06/01/1002196245/a-u-n-report-suggests-libya-saw-the-first-battlefield-killing-by-an-autonomous-d ("[A] United Nations report about a March 2020 skirmish in the military conflict in Libya says such a drone, known as a lethal autonomous weapons system — or LAWS — has made its wartime debut."); *see also* The Editorial Board, *Rules of War Need Rewriting for the Age of AI Weapons*, FIN. TIMES (Dec. 1, 2021), https://www.ft.com/content/d8371144-364b-496d-943c-16f7e0982b6e ("The UN says Turkish-made Kargu drones incorporating image-processing capabilities were used in Libyan conflicts last year to home in on selected targets.").

In addition, even before AI becomes a significant component of battlefield engagement, AI will be part of the cyberwarfare that will accompany any battle.[72] Computer systems are increasingly at the heart of every information system and weapon system. Although cybersecurity measures will continue to improve,[73] history suggests that the ability to attack complex systems grows more rapidly than the ability to protect them.[74] "AI and machine learning can help to keep abreast with cybercriminals, automate threat detection, and respond more effectively than conventional software-driven or manual techniques."[75] Unfortunately, an attacker can employ AI to maximize exploits in precisely the same manner a cybersecurity expert can use it to thwart incursions. "[C]ybercriminals can also use AI to analyze their malware and launch more advanced attacks . . . ."[76]

Worse, AI is itself vulnerable to attacks using weaponized data or other means of manipulating the AI systems "in order to alter their behavior to serve a malicious end goal."[77] These attacks can be directed at a combination of military targets and civilian targets. "There are five areas most immediately affected by artificial intelligence attacks: content filters, the military, law enforcement, traditionally human-based tasks being replaced by AI, and civil society."[78] As AI grows in importance in each of these sectors, the risk of attack increases as well.[79]

As AI systems become indispensable, they become points of failure for military and civilian systems. "AI is expanding the window of vulnerability the United States has already entered. For the first time since World War II, America's technological predominance—the backbone of its economic and military power—is under threat."[80] In late 2021, thousands of businesses were forced to shut down

---

72. The Editorial Board, *supra* note 71 ("Beyond killer robots, AI could be used to enhance or replace human skill in everything from operating weapons to intelligence gathering and analysis, early warning systems, and command and control.").

73. Michael E. O'Hanlon, *The Role of AI in Future Warfare*, BROOKINGS INST. (Nov. 29, 2018), https://www.brookings.edu/research/ai-and-future-warfare/ ("By 2040, many cyber systems controlling NATO weaponry and other platforms should be more resilient to attack. That is because NATO will have had two decades to address problems that are now widely understood.").

74. THE PRESIDENT'S NAT'L INFRASTRUCTURE ADVISORY COUNCIL, SECURING CYBER ASSETS: ADDRESSING URGENT CYBER THREATS TO CRITICAL INFRASTRUCTURE 7 (Aug. 2017) https://www.cisa.gov/sites/default/files/publications/niac-securing-cyber-assets-final-report-508.pdf ("The scale, scope, and frequency of cyber attacks on digital and physical infrastructure systems is growing rapidly. Threats are escalating as more sophisticated and organized attackers are designing targeted attacks to damage or disrupt vital services and critical physical systems.")

75. Gaurav Belani, *The Use of Artificial Intelligence in Cybersecurity: A Review*, IEEE COMPUTER SOCIETY (last visited Dec. 23, 2021), https://www.computer.org/publications/tech-news/trends/the-use-of-artificial-intelligence-in-cybersecurity.

76. *Id.*

77. Marcus Comiter, *Attacking Artificial Intelligence: AI's Security Vulnerability and What Policymakers Can Do About It*, BELFER CTR. FOR SCI. AND INT'L AFFS., HARV. KENNEDY SCHOOL (Aug. 2019), https://www.belfercenter.org/publication/AttackingAI.

78. *Id.*

79. *Id.*

80. ERIC SCHMIDT ET AL., FINAL REPORT: NATIONAL SECURITY COMMISSION ON ARTIFICIAL INTELLIGENCE 7 (Mar. 19, 2021) [hereinafter NSCAI FINAL REPORT], https://www.nscai.gov/wp-content/uploads/2021/03/Full-Report-Digital-1.pdf

when Facebook's platform or Amazon's AWS suffered an outage.[81] If these were targeted as part of a cyberwarfare campaign, the potential exists for even greater vulnerability.

The threat cannot be understated:

> Because of AI, adversaries will be able to act with micro-precision, but at macro-scale and with greater speed. They will use AI to enhance cyber attacks and digital disinformation campaigns and to target individuals in new ways. AI will also help create precisely engineered biological agents. And adversaries will manipulate the AI systems we will rely upon.[82]

The use of AI and the other techniques described in this Article create a facile toolbox for adversaries seeking to destabilize regimes and benefit from a weakened and internally conflicted United States. "[T]hese AI-enabled capabilities will be used across the spectrum of conflict. They will be used as tools of first resort in non-military conflicts, as a prelude to military actions, or in concert with military actions in war."[83]

Beyond the power of AI, concern also exists because AI lacks the common sense or gut instinct to know when it is being provided false data. The 1983 film War Games[84] captures the fear that a smart computer system may not know that the data suggesting it was defending against a nuclear attack was merely a simulation. AI systems have glaring blind spots. "Researchers have already demonstrated how to fool an AI system into misreading a stop sign, by carefully positioning stickers on it. They have deceived facial-recognition systems by sticking a printed pattern on glasses or hats. And they have tricked speech-recognition systems into hearing phantom phrases . . . ."[85] In a war-like setting, AI has a host of potential vulnerabilities, which include the external manipulation of the information being observed but also extends to hacking the systems on which

---

81. *See* Richard Lawler, *Facebook Explains the Backbone Shutdown Behind its Global Outage on Monday*, THE VERGE (Oct 5, 2021, 2:29 PM), https://www.theverge.com/2021/10/5/22710963/facebook-dns-bgp-outage-backbone-maintenance (a combination of an accidental shutdown of the backbone and then the DNS servers followed programming to eliminate routing even to the restored systems); *see also* Jeremy C. Fox, *Amazon Web Services Outage Leads to Cancellation of Parking Tickets, Shuts Down Local Services in Boston*, BOSTON GLOBE (Dec. 7, 2021, 8:06 PM), https://www.bostonglobe.com/2021/12/07/business/amazon-network-services-outage-shuts-down-boston-parking-app-other-local-services/ ("Amazon Web Services, which provides cloud computing services to many companies, governments, and universities, experienced a major outage on Tuesday that disrupted access to websites and electronic services.").

82. NSCAI FINAL REPORT, *supra* note 80, at 45 ("state and non-state adversaries are challenging the United States below the threshold of direct military confrontation by using cyber attacks, espionage, psychological and political warfare, and financial instruments.").

83. *Id.* at 46.

84. WAR GAMES (United Artists & Sherwood Productions 1983).

85. Douglas Heaven, *Why Deep-Learning AIs are so Easy to Fool*, NATURE (Oct. 9, 2019), https://www.nature.com/articles/d41586-019-03013-5.

the AI relies.[86] Using these techniques, an aggressor does not need to build the world-ending AI system; the aggressor merely needs to trick the system into a response mode that could launch a conflict of regional or even global scale.

B. Synthetic Media, Deepfakes, and Other Falsification Techniques

Artificial Intelligence systems are so pervasive that the concerns about their use and misuse may be analogized to similar concerns over electricity or gasoline engines. They are ubiquitous technologies essential to many socially-valuable services. Although they are capable of being put to great harm, that is equally true of all inventions. After all, the Nobel Peace Prize was funded by dynamite inventor Alfred Nobel, who created an explosive designed for mining which was quickly exploited for military aims.[87] Propaganda, however, is another story. "Propaganda, particularly through the medium of the radio, becomes a grave menace to peace when used by an aggressive state to stir up hatred, revolution, and war."[88] The radio of World War II has been replaced by mass media, social media, and virtual worlds, making its potency greater than ever.

Media falsification tools such as deepfake software have only modest beneficial use[89] and a much larger potential for misinformation, propaganda, and public deception. "'[S]ynthetic media'—more commonly known as 'deepfakes' ... are hyper-realistic video and audio recordings that use artificial intelligence and 'deep' learning to create 'fake' content or 'deepfakes.' The U.S. government has grown increasingly concerned about their potential to be used to spread disinformation and commit crimes."[90] Synthetic media may be digitally modified,

---

86. *See* Christian Berghoff, Matthias Neu & Arndt von Twickel, *Vulnerabilities of Connectionist AI Applications: Evaluation and Defense*, FRONTIERS IN BIG DATA (July 22, 2020), https://doi.org/10.3389/fdata.2020.00023 ("Possible threats include augmenting the training data set with poisoned data to sabotage training, changing the hyperparameters of the training algorithm or directly changing the model's parameters . . . . Furthermore, an attacker may manipulate already trained models . . . . by retraining the models with specially crafted data in order to insert backdoors . . . .").

87. *See* Sven Tägil, *Alfred Nobel's Thoughts about War and Peace,* NOBEL PRIZE (last visited Dec. 23, 2021), https://www.nobelprize.org/alfred-nobel/alfred-nobels-thoughts-about-war-and-peace/ ("[Alfred Nobel's] great invention, dynamite, had not been developed with the idea of using it in war. However, this did not prevent it from soon being put to use in such a context as well. Dynamite was used, for example, in the Franco-Prussian War first by the Prussians, and later also by the French.").

88. John B. Whitton, *Efforts to Curb Dangerous Propaganda*, 41 AM. J. INT'L LAW 899, 899 (1947) (citing John B. Whitton & John H. Herz, *The Radio in International Politics, in* PROPAGANDA BY SHORT-WAVE Chapter 1 (Princeton Univ. Press & London Oxford Univ. Press 1942)).

89. *See* James Vincent, *Disney's Deepfakes are Getting Closer to a Big-Screen Debut*, THE VERGE (June 29, 2020, 11:53 AM), https://www.theverge.com/2020/6/29/21306889/disney-deepfake-face-swapping-research-megapixel-resolution-film-tv.

90. Bill Whitaker, *Synthetic Media: How Deepfakes Could Soon Change Our World*, CBS NEWS (Oct. 10, 2021, 6:54 PM) https://www.cbsnews.com/news/deepfake-artificial-intelligence-60-minutes-2021-10-10/.

or it may be wholly generated by AI.[91] "Deepfakes use deep learning artificial intelligence to replace the likeness of one person with another in digital media.[92] With the rise of deepfakes and synthetic media, adversaries are leveraging this technology to create fake news and misleading, counterfeit videos."[93]

Russia uses disinformation and propaganda aggressively to maintain its position in the world. "The Kremlin aims to leverage shared elements of the post-Soviet experience in order to drive wedges between ethnic Russian or Russian-speaking populations who reside in these states and their host governments."[94] The purpose is not just to slowly reclaim authority for the power formerly held by the Soviet Union. "Farther abroad, the Kremlin attempts to achieve policy paralysis by sowing confusion, stoking fears, and eroding trust in Western and democratic institutions."[95] "AI is deepening the threat posed by cyber attacks and disinformation campaigns that Russia, China, and others are using to infiltrate our society, steal our data, and interfere in our democracy. The limited uses of AI-enabled attacks to date represent the tip of the iceberg."[96]

Russia has already used synthetic media to augment its disinformation campaign during its 2014 invasion of Crimea and Eastern Ukraine. Russia is using similar tactics in its 2021 campaign to further destabilize Ukraine. The Russian government-sponsored RT (formerly Russia Today) international television network was one component of the campaign.[97] "RT's editor-in-chief, Margarita Simonyan, gave an interview in 2014 in which she said that RT was 'fighting' for Russia by 'conducting the information war' against 'the whole of the Western world.'"[98] Russian-backed separatists used a Russian-supplied missile to shoot

---

91. NINA SCHICK, DEEPFAKES: THE COMING INFOCALYPSE 5 (Grand Cent. Publ'g 2020) (noting that infocalypse "was coined by the U.S. technologist Aviv Ovadya in 2016").

92. The U.S. Department of State commented:

> As the U.S. Government's dedicated center for countering foreign disinformation and propaganda, the Global Engagement Center (GEC) at the U.S. Department of State has a mandate to expose and counter threats from malign actors that utilize these tactics. In this field, Russia continues to be a leading threat. The Department works with interagency and global partners to meet this challenge, with the GEC playing a key role in coordinating efforts and helping lead a global response.

See U.S. DEPT. OF STATE, GEC SPECIAL REPORT: PILLARS OF RUSSIA'S DISINFORMATION AND PROPAGANDA ECOSYSTEM 3 (Aug. 2020), https://www.state.gov/wp-content/uploads/2020/08/Pillars-of-Russia%E2%80%99s-Disinformation-and-Propaganda-Ecosystem_08-04-20.pdf.

93. Sarah Sybert, *DARPA Launches New Programs to Detect Falsified Media*, GOVERNMENT CIO MEDIA & RESEARCH (Sept. 16, 2021), https://governmentciomedia.com/darpa-launches-new-programs-detect-falsified-media.

94. TODD C. HELMUS ET AL., RUSSIAN SOCIAL MEDIA INFLUENCE: UNDERSTANDING RUSSIAN PROPAGANDA IN EASTERN EUROPE x (RAND Corp. 2018), https://www.rand.org/content/dam/rand/pubs/research_reports/RR2200/RR2237/RAND_RR2237.pdf

95. *Id.*

96. NSCAI FINAL REPORT, *supra* note 80, at 7.

97. SCHICK, *supra* note 91, at 8-10.

98. *Id* at 11.

down a civilian aircraft, and then Russia aggressively campaigned to pin the blame on the Ukrainians.[99] Russia went even further. "Russia simply denied that it had invaded Ukraine, claiming that western politicians and commentators were waging an unjustified anti-Russia smear campaign."[100]

It is unclear to say how extensively Russia already uses synthetic media to bolster its disinformation campaigns, but the evidence already establishes that the campaigns are ongoing.[101] Researchers "found a sprawling web of nonexistent authors turning Russian-government talking points into thousands of opinion pieces and placing them in sympathetic Western publications, with crowds of fake people discussing the same themes on Twitter."[102] A subsequent campaign replaced stolen photographs and faked journalistic credentials with AI-generated images and wholly fictionalized backgrounds that made detection even harder.[103]

Synthetic media is not the only technique for widescale public manipulation. As already noted, Russia's government-owned RT television network produces content often found false and misleading.[104] It is used to delegitimize democratic regimes and justify the continued rule by its current regime. Russia has used its disinformation platform to attack the political process in the U.S. and Europe.[105]

---

99. Gerard Toal, John O'Loughlin & Kristin M. Bakke, *Flight MH17 Crashed Six Years Ago. Ukrainians Have Very Different Views on Who's to Blame.*, WASH. POST (July 16, 2020), https://www.washingtonpost.com/politics/2020/07/16/flight-mh17-crashed-six-years-ago-ukrainians-have-very-different-views-whos-blame/ ("In October 2015 the joint investigative team concluded a Buk surface-to-air missile launched from pro-Russian separatist-controlled territory in Ukraine downed the aircraft. . . . Direct Russian military intervention supported separatists in Ukraine.").

100. SCHICK, *supra* note 91, at 8.

101. *See* Renée DiResta, *The Supply of Disinformation Will Soon Be Infinite*, THE ATLANTIC (Sept. 20, 2020), https://www.theatlantic.com/ideas/archive/2020/09/future-propaganda-will-be-computer-generated/616400/.

102. *Id.*

103. *Id.*

104. *See* Robert Elliott, *How Russia Spreads Disinformation via RT is More Nuanced than we Realise*, GUARDIAN (July 26, 2019), https://www.theguardian.com/commentisfree/2019/jul/26/russia-disinformation-rt-nuanced-online-ofcom-fine ("A detailed analysis of the output of state-controlled media often reveals a fundamental disparity between Russian-owned outlets in the UK and the rest of Europe when comparing them to established news organisations. The volume of coverage, framing of coverage, and average engagement with that coverage is, at times, widely disparate."); *see also* GORDON RAMSAY & SAM ROBERTSHAW, WEAPONISING NEWS: RT, SPUTNIK AND TARGETED DISINFORMATION 44 (King's College London Jan. 2019), https://www.kcl.ac.uk/policy-institute/assets/weaponising-news.pdf ("T]he consistent portrayal of Western governments as untrustworthy partners . . . driven by dishonest or hypocritical goals and both dangerously aggressive and simultaneously weak and unstable, suggests that the emergence of these groups of narratives may have been a conscious editorial strategy.").

105. *See* Nomaan Merchant, *Congress Ordered up a Nerve Center to Stop Election Interference in 2019. It Still Doesn't Exist*, L.A. TIMES (Dec. 26, 2021, 4:38 AM), https://www.latimes.com/world-nation/story/2021-12-23/congress-ordered-up-a-nerve-center-to-stop-election-interference-in-2019-it-still-doesnt-exist.

> U.S. and other Western authorities have accused Russia of spreading disinformation about the coronavirus and vaccines, stealing data from local and

Russia hacked the networks of the U.S. Democratic Party, released thousands of emails, falsified social media accounts and more.[106] In Europe, Russia has funneled funds to political parties, created artificial political movements, and "even supported a failed coup in Montenegro to unseat the pro-NATO government."[107] To promote its disinformation, Russia has also invested in "'troll farms'—groups of organized online agitators—identify grievances in other countries and then insert themselves into those debates with the aim of inflaming them."[108]

So far, the U.S. response has been tepid. "As Russia was working to subvert U.S. elections and sow discord among Americans, Congress directed the creation of an intelligence center to lead efforts to stop interference by foreign adversaries. But two years later, that center still is not close to opening."[109] Adversaries that tried to interfere in the last two presidential elections continue to bombard Americans with disinformation and conspiracy theories at a time of peril for democracy in the U.S. and around the world.

Russia is not alone. China has an even larger program for online disinformation. China leans heavily on the public using paid workers, often government employees, to post hundreds of millions of false or misleading social media comments annually.[110]

Evidence suggests that many of these posts are created by "regular government employees . . . that . . . work directly for the Communist Party or for different organs of the local government, and presumably are expected to write these comments as part of their official duties."[111] According to research conducted by Gary King, Jennifer Pan and Margaret Roberts under a National Science

---

state election servers and pushing false stories intended to exploit divisions over race and civil rights. Intelligence agencies have found that Russia used influence operations to interfere with the 2016 presidential election in favor of Trump's campaign and conducted operations in Trump's favor in 2020.

106. Maggie Tennis, *Russia Ramps up Global Elections Interference: Lessons for the United States*, CENTER FOR STRATEGIC AND INT'L STUDIES (July 20, 2020), https://www.csis.org/blogs/technology-policy-blog/russia-ramps-global-elections-interference-lessons-united-states.

107. *Id.*

108. Scottie Barsotti, *Weaponizing Social Media: Heinz Experts on Troll Farms and Fake News*, CARNEGIE MELLON UNIV. HEINZ COLLEGE (last visited Dec. 26, 2021), https://www.heinz.cmu.edu/media/2018/October/troll-farms-and-fake-news-social-media-weaponization ("Rather than promoting any one political ideology, professional Russian trolls instead focus on fanning Americans' emotions around heated topics such as gun control or immigration, and then pitting Americans against Americans. The tactic is—literally—divide and conquer.")

109. Merchant, *supra* note 105.

110. Henry Farrell, *The Chinese Government Fakes Nearly 450 Million Social Media Comments a Year. This is Why.*, WASH. POST (May 19, 2016), https://www.washingtonpost.com/news/monkey-cage/wp/2016/05/19/the-chinese-government-fakes-nearly-450-million-social-media-comments-a-year-this-is-why/ ("Internet researchers have long known that the Chinese government manipulates content on the Internet. Not only does it censor heavily, but it also employs hundreds of thousands of people, the so-called 50 cent army, to write comments on the Internet.").

111. *Id.*

Foundation grant,[112] "the strategic objective of the regime is to distract and redirect public attention from discussions or events with collective action potential."[113] This usage differs somewhat from the Russian disinformation model, but it is very consistent with China's focus on suppressing dissent where such dissent could lead to civil unrest.[114]

Social media comments can be merely text. But more sophisticated campaigns involve doctored images and edited or altered videos. Photography and videography, of course, were never free of manipulation and editorial framing. "[D]igital imaging has simply forced everyone to acknowledge the inherently manipulative nature of photography and to understand that it never represented 'truth' in the first place."[115] In the days of analog photography, however, there were often mismatched shadows, stitch lines, and other visual clues to media manipulation, but the creation of synthetic media will often avoid such obvious signs.[116]

Synthetic media has the ability to produce any combination of wholly artificial still imagery, videography, audio, and text, which makes it inherently more manipulative, and to produce this exceedingly manipulative content at an unprecedented scale.[117] "Deepfake technology can generate, for example, a humorous, pornographic, or political video of a person saying anything, without the consent of the person whose image and voice is involved."[118] Deepfakes and other forms of synthetic media undermine trust in media outlets and encourage the public to rely on social media, which is particularly vulnerable to synthetic media.[119]

---

112. Gary King, Jennifer Pan & Margaret E. Roberts, *How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument*, 111 AM. POL. SCI. REV. 484, 484 (2017).

113. *Id.* at 485.

114. *See* Louisa Lim, *In China, A Ceaseless Quest To Silence Dissent*, NAT. PUB. RADIO (Oct. 30, 2012, 12:36 PM), https://www.npr.org/2012/10/30/163658996/in-china-a-ceaseless-quest-to-silence-dissent.

115. *See* Leslie Mullen, Truth In Photography: Perception, Myth And Reality In The Postmodern World (1988) U. FL. DOCTORAL DISSERTATION 10 (1998) (M.A. thesis, University of Florida), https://docplayer.net/10735211-Truth-in-photography-perception-myth-and-reality-in-the-postmodern-world.html ("So a photograph, although professing to depict truth, actually involves manipulation of both object and message. . . . The photographer chooses what aspect of reality he wishes to represent both when he takes the picture, and when he readies it for publication."), *id.* at 8-9.

116. *See id.*

117. *See* CHRISTINA NEMR & WILLIAM GANGWARE, WEAPONS OF MASS DISTRACTION: FOREIGN STATE-SPONSORED DISINFORMATION IN THE DIGITAL AGE 40-41 (Park Advisors Mar. 2019), https://www.state.gov/wp-content/uploads/2019/05/Weapons-of-Mass-Distraction-Foreign-State-Sponsored-Disinformation-in-the-Digital-Age.pdf ("The sheer magnitude of content and platforms is perhaps one of the biggest obstacles hindering monitoring and detection. . . . [T]he enormous and steadily growing volume of content being uploaded raises questions about the ability of these platforms to effectively monitor all of it.")

118. Mika Westerlund, *The Emergence of Deepfake Technology: A Review*, 9 TECH. INNOVATION MGMT. REV. 39, 39 (Nov. 2019) (internal citations omitted).

119. *See id.*

> Deepfakes target social media platforms, where conspiracies, rumors, and misinformation spread easily, as users tend to go with the crowd. At the same time, an ongoing 'infopocalypse' pushes people to think they cannot trust any information unless it comes from their social networks, including family members, close friends or relatives, and supports the opinions they already hold. . . . Deepfakes are a major threat to our society, political system, and business because they 1) put pressure on journalists struggling to filter real from fake news, 2) threaten national security by disseminating propaganda and interfering in elections, 3) hamper citizen trust toward information by authorities, and, 4) raise cybersecurity issues for people and organizations.[120]

Political manipulation using synthetic media has already begun.

> In May [2018], a video appeared on the internet of Donald Trump offering advice to the people of Belgium on the issue of climate change. "As you know, I had the balls to withdraw from the Paris climate agreement," he said, looking directly into the camera, "and so should you." The video was created by a Belgian political party, Socialistische Partij Anders, or sp.a, and posted on sp.a's Twitter and Facebook. It provoked hundreds of comments, many expressing outrage that the American president would dare weigh in on Belgium's climate policy. . . . The speech, it was later revealed, was nothing more than a hi-tech forgery.[121]

Although the politicians who exploited the deepfake claimed the poor quality should have alerted the public to the obvious parody, the political party itself did nothing to warn the viewers or contextualize the message.[122]

In China, the state news agency has used synthetic media to create AI-based digital news anchors who provide lifelike media coverage. "Not only can I accompany you 24 hours a day, 365 days a year. I can be endlessly copied and present at different scenes to bring you the news," explains the digital version of Xinhua news anchor Qiu Hao.[123] The combination of synthetic, government-sponsored news anchors and falsified video will make it increasingly difficult for the public to exercise effective digital literacy. As a result, the infopocalypse will intensify and distrust will expand even further.

---

120. *Id.* at 39, 42 (citations omitted).

121. Oscar Schwartz, *You Thought Fake News was Bad? Deep Fakes are Where Truth Goes to Die*, GUARDIAN (Nov. 12, 2018), https://www.theguardian.com/technology/2018/nov/12/deep-fakes-fake-news-truth.

122. *See id.*

123. Lily Kuo, *World's First AI News Anchor Unveiled in China*, GUARDIAN (Nov. 8, 2018), https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china.

There are many efforts to combat the rapid growth of synthetic media.[124] DARPA has launched a program to identify tell-tale signs of digital manipulation.[125] Facebook and Michigan State have a system that is approximately 70 percent accurate, the most reliable system as of the time of this writing.[126]

More sophisticated systems are needed both because of the sophistication of the fakes and because of the scale of the deployment. A system that is 70-80 percent effective has the potential to miss thousands—or even millions—of fake posts. Add the disinformation from state media, campaigns by bots and paid employees to reinforce the misinformation in the false videos, and the trustworthiness of media will become increasingly eroded. "Bad actors who understand those mechanisms and user tendencies have used that knowledge to weaponize information in various ways, such as swaying public opinion or sowing chaos in the leadup to an election."[127]

Without taking much more aggressive steps, the use of synthetic media and other disinformation techniques will continue to erode democratic engagement, making it easier and easier for foreign adversaries and non-democratic influences to gain power over governments in the U.S. and abroad. Unless there is a new system of accountability, synthetic media will ultimately eat away at all democratic institutions.

C. Virtual Worlds and the Metaverse

Synthetic media and related disinformation techniques are a pervasive problem for a free society. Fueled by continuous improvements in AI, the impact of these techniques will continue to expand each year. In coming years, however, the

---

124. *See* Matt Groh, *Detect DeepFakes: How to Counteract Misinformation Created by AI*, MIT MEDIA LAB (last visited Dec. 26, 2021), https://www.media.mit.edu/projects/detect-fakes/overview/; *see also* Leo Kelion, *Deepfake Detection Tool Unveiled by Microsoft*, BBC NEWS (Sept. 1, 2020), https://www.bbc.com/news/technology-53984114 ("Microsoft's Video Authenticator tool works by trying to detect giveaway signs that an image has been artificially generated, which might be invisible to the human eye. These include subtle fading or greyscale pixels at the boundary of where the computer-created version of the target's face has been merged with that of the original subject's body."); *see also* Jeremy Kahn, *Facebook Says it's Made a Big Leap Forward in Detecting Deepfakes*, FORTUNE (June 16, 2021), https://fortune.com/2021/06/16/facebook-detecting-deepfakes-research-michigan-state/ (A "research team from Facebook and Michigan State say they have created a system that, at 70% accuracy on a key benchmark test, is significantly better than any previous system that ingested whole still images or video frames for examination.).

125. Sarah Sybert noted:

> DARPA's Media Forensics program (MediFor) builds algorithms to detect manipulated images or videos, then produces a quantitative measure of integrity, which enables filtering and prioritization of media at scale. The agency is focusing on three types of integrity: digital, physical and semantic. MediFor uses detection algorithms, which analyze media content to determine if manipulation has occurred, and fusion algorithms, which combine information across multiple detectors to create a unified score for each media asset

Sybert, *supra* note 93.

126. Kahn, *supra* note 124.

127. Barsotti, *supra* note 108.

potential for their impact will be even greater than any time in history. This is because the growth of the metaverse will legitimize the use of artificial media for many Americans' daily lives. When this happens, the tools to separate disinformation from reality will be lost to the public.

The metaverse, as used in this context, is a collection of platforms that share a similar user interface which relies on virtual reality or augmented reality designed to facilitate the interaction between participants and to provide interactions among participants with computer-generated characters.[128] Ideally, each metaverse host will enable the digital assets and avatars on that platform to be persistent and interoperable with other metaverses on other platforms.[129] The metaverse "also translates to a digital economy, where users can create, buy, and sell goods. And, in the more idealistic visions of the metaverse, it's interoperable, allowing you to take virtual items like clothes or cars from one platform to another."[130]

Like AI, there are a multitude of uses for virtual worlds and the metaverse or metaverses. Among the two most significant metaverses today are the world-building platform Roblox and the virtual world game Fortnite.[131] Roblox, in particular, is considered the originator of the current land rush into virtual world investment, triggered by the company's initial public offering in early 2021.[132]

---

128. Eric Ravenscraft noted:

> [T]he term [metaverse] doesn't really refer to any one specific type of technology, but rather a broad shift in how we interact with technology. . . . [T]he technologies that make up the metaverse can include virtual reality . . . as well as augmented reality that combines aspects of the digital and physical worlds.

See Eric Ravenscraft, *What Is the Metaverse, Exactly?*, WIRED (Nov. 25, 2021, 7:00 AM), https://www.wired.com/story/what-is-the-metaverse/.

129. *See* Przemyslaw Pałka, *The World of Fifty (Interoperable) Facebooks*, 51 SETON HALL L. REV. 1193, 1229–30 (2021) ("Put simply, products are interoperable if they can work together. . . . 'the ability to transfer and render useful data and other information across systems, applications, or components.' [John Palfrey and Urs Gasser] nuance the definition by distinguishing four layers of interoperability: technological, data, human, and institutional.") (quoting JOHN PALFREY & URS GASSER, INTEROP: THE PROMISE AND PERILS OF HIGHLY INTERCONNECTED SYSTEMS 5 (Basic Books 2012).

130. Ravenscraft, *supra* note 128.

131. *See* Nick Statt, *Fortnite Inches Closer to the Metaverse with New Party Worlds*, PROTOCOL (Dec. 1, 2021), https://www.protocol.com/bulletins/fortnite-party-worlds-metaverse-epic?share_id=6808523 ("Epic is making its hit game Fortnite less about firearms and more about self-expression and socializing. It launched a new game format that looks like it inches Epic closer to the sought-after metaverse so many tech and gaming firms keep going on about.").

132. *See* ERIC SHERIDAN ET AL., FRAMING THE FUTURE OF WEB 3.0: METAVERSE EDITION 4 (Goldman Sachs Group, Inc. 2021), https://www.goldmansachs.com/insights/pages/gs-research/framing-the-future-of-web-3.0-metaverse-edition/report.pdf ("Over the past 12 months, the term Metaverse began to gain traction shortly after Roblox's direct listing in March and more meaningfully saw higher levels of Google Search interest during the Q3 '21 earnings season as various management teams discussed elements of their business within the future Metaverse."); *see also* Dean Takahashi, *The DeanBeat: Roblox Public Offering is a Vote About the Metaverse*, GAMESBEAT (Mar. 5, 2021, 8:00 AM), https://venturebeat.com/2021/03/05/the-deanbeat-roblox-public-offering-is-a-vote-about-the-metaverse/ ("Roblox, the platform for user-generated games, will go public through a direct listing of its shares on March 10. I see its pending success or failure

Fortnite is transforming some of its experiences into social engagements. "Epic ... says these virtual spaces 'should have a high focus on self-expression through emotes, sprays, outfit changes, or other mechanics,' and that they should 'encourage social interaction, giving people a way to make new friends or team up with existing friends in new ways.'"[133]

The metaverse or the metaverses have the potential to be used in a wide variety of ways beyond gaming. Companies such as Microsoft are looking to make aspects of the metaverse an extension of the work-from-anywhere ethos. Mesh for Teams, a new Microsoft service, "allows workers to take the form of avatars and navigate virtual work environments. It combines 'shared holographic experiences' with existing communication tools like virtual meetings, chats and shared documents."[134] Others are exploring the potential to operate courses or schools utilizing the metaverse.[135] And, of course, business and ecommerce will be augmented through metaverse interactions.[136]

Although the metaverse remains in its nascent state, it will build on the current Web 2.0 internet. Venture Capitalist analyst and investor Matthew Ball has offered seven attributes that describe the metaverse as well as the current internet: persistence; synchronous and live interactions; the capacity for as many concurrent users as the users demand; a stable, functioning economy; the incorporation of both digital and physical worlds as well as operating on both open and closed platforms; largely interoperable; and be "populated by 'content' and 'experiences' created and operated by an incredibly wide range of contributors."[137]

---

as a stock as a kind of referendum on the metaverse. . . ."); *see also* Patrick Seitz, *Roblox Stock Continues Meteoric Rise On Metaverse Story*, INVESTOR'S BUS. DAILY (Nov. 19, 2021), https://www.investors.com/news/technology/roblox-stock-continues-meteoric-rise-on-metaverse-story/?fbclid=IwAR1zi2ocDxFQzGYNaWD0nDR0YFcANegqUTNy-VT_3zJUsCP5KuMnCWGsUvE ("Roblox today provides a platform for playing video games and socializing in 3D virtual worlds. But Roblox stock is considered a play on the metaverse, a next-generation version of the internet.").

133. Statt, *supra* note 131; *see also* Kirk McKeand, *Epic Games, the Metaverse and the Terrifying Consolidation of the Games Industry*, USA TODAY (Nov. 24, 2021, 10:17 AM), https://ftw.usatoday.com/2021/11/epic-games-fortnite-metaverse-big-tech.

134. Tony Lystra, *Microsoft Offers its Own Take on the Metaverse with the Introduction of Mesh for Teams*, GEEKWIRE (Nov. 2, 2021, 8:00 AM), https://www.geekwire.com/2021/microsoft-offers-take-metaverse-introduction-mesh-teams/ (quoting Microsoft).

135. *See* Ellysse Dick, *The Promise of Immersive Learning: Augmented and Virtual Reality's Potential in Education*, INFORMATION TECHNOLOGY & INNOVATION FOUND. (Aug. 30, 2021), https://itif.org/publications/2021/08/30/promise-immersive-learning-augmented-and-virtual-reality-potential; Kwang Hyung Lee, *The Educational 'Metaverse' is Coming*, THE CAMPUS (Oct. 29, 2021) https://www.timeshighereducation.com/campus/educational-metaverse-coming ("The time has come to rebuild the curriculum and infrastructure for the world of the metaverse. We can't go back to the way things were before.").

136. *See* Beth Owens, *Ecommerce and the Metaverse: What we can Expect*, WHIPLASH (Dec. 7, 2021), https://whiplash.com/blog/ecommerce-and-the-metaverse/.

137. Matthew Ball, *The Metaverse: What it is, Where to Find it, and Who Will Build it*, MATTHEWBALL.VC (Jan 13, 2020), https://www.matthewball.vc/all/themetaverse; *see also* Ben Thompson, *Microsoft and the Metaverse*, STRATECHERY (Nov. 9, 2021),

The attributes of the metaverse make it ideal to augment work, education, socialization, prayer, and entertainment. In other words, the metaverse has the potential to be as disruptive to the world as the internet on which it is being built.[138] The metaverse is being embraced by many of the leading tech companies as well as by advocates for a much more decentralized internet economy.[139] Known as Web3, the decentralized model of the future internet is built on blockchain technology rather than corporate data as the basis for user information persistence.[140] "In a Web3 world, people control their own data and bounce around from social media to email to shopping using a single personalized account, creating a public record on the blockchain of all of that activity."[141] The proposed Web3 infrastructure relies on a peer-to-peer sharing system verified using blockchain private-key security.[142] This reduces the potential points of failure triggered when large providers have outages.[143]

The large tech companies that dominate the current Web 2.0 infrastructure will not disappear. The most likely scenario is that Web3 experiences will be additive to the existing internet. Particular competitors may emerge, and popular sites may falter, just as Yahoo! and MySpace went from dominating the internet to being niche players.

Web3 is a slightly larger arena of new services that incorporates technologies such as the metaverse, fintech innovations in cryptocurrency, and digital assets represented (or in the form of) nonfungible tokens (NFTs).[144] Each of these innovations has a wide range of legitimate uses. Like other technologies, they also can be used to support foreign military, terrorist, or criminal activities.

Cryptocurrencies, in particular, have raised considerable concerns regarding their usefulness in supporting criminal endeavors, facilitating ransomware, assisting money laundering, and hiding payments by belligerent nations to troll

---

https://stratechery.com/2021/microsoft-and-the-metaverse/ (quoting Ball and noting that these attributes describe the internet as well).

138. *See* Jon M. Garon, *Legal Implications of a Ubiquitous Metaverse and a Web3 Future*, ABA CYBERSPACE COMM. WINTER WORKING MEETING (Jan. 19, 2022), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4002551.

139. *See* Bobby Allyn, People are talking about Web3. Is it the Internet of the Future or Just a Buzzword?, Nat. Pub. Radio (Nov. 21, 2021), https://www.npr.org/2021/11/21/1056988346/web3-internet-jargon-or-future-vision ("Platforms like Google, Amazon, Facebook and Twitter emerged to bring order to the Internet by making it easy to connect and transact online. Critics say over time those companies amassed too much power. Web3 is about grabbing some of the power back.").

140. *Id.*; *see also* Thibault Meunier & In-Young Jo, *Web3 — A Vision for a Decentralized Web*, Cloudflare (Oct. 1, 2021), https://blog.cloudflare.com/what-is-web3/.

141. Allyn, *supra* note 139.

142. *See* Meunier & Jo, *supra* note 140.

143. *See id.*

144. *See* Gilad Edelman *The Father of Web3 Wants You to Trust Less*, WIRED (Nov. 29, 2021), https://www.wired.com/story/web3-gavin-wood-interview/ ("At the most basic level, Web3 refers to a decentralized online ecosystem based on the blockchain. Platforms and apps built on Web3 won't be owned by a central gatekeeper, but rather by users, who will earn their ownership stake by helping to develop and maintain those services.").

farms and terror cells.[145] Traditional financial service institutions are in a position to "observe suspicious activity" and help the federal government intervene in these activities.[146]

Unlike cryptocurrencies, the potential threat posed by the metaverse is much less well defined or identified.[147] For foreign governments interested in disinformation and the destabilization of their enemies as well as for terrorist groups looking to radicalize followers, launder money, or attack avowed enemies, the synthetic nature of the metaverse makes it easier than ever to hide the identity of the person on the other side of an interaction or to create AI-based avatars and bots to interact and promote the foreign agent's agenda.

China, in particular, has indicated that it views dominance in the metaverse as essential to its national strategy.

> China . . . launched its first metaverse industry group: the Metaverse Industry Committee, under the state-supervised China Mobile Communications Association (CMCA). Speaking at the launch ceremony, . . . the former vice minister of the ministry of science and technology Wu Zhongze laid out the high stakes of the incipient metaverse. He made clear that it was no passing fad nor empty buzzword, but rather an important trend to seize on as China seeks to cement its global technological prowess.[148]

China has taken considerable steps to anticipate the growth of the metaverse, incorporating it into the strategy of increased state control of data and digital assets. China also has the lead over the U.S. in AI investment, consumer product manufacturing, and online gaming, giving it the potential to dictate at least some of the norms that are established in the metaverse.[149] "The [Chinese] government

---

145. *See* Stan Sater, *Do We Need Kyc/aml: The Bank Secrecy Act and Virtual Currency Exchanges*, 73 Ark. L. Rev. 397, 423 (2020); CYBER-DIGITAL TASK FORCE, UNITED STATES DEPARTMENT OF JUSTICE OFFICE OF THE DEPUTY ATTORNEY GENERAL (October 2020), https://www.justice.gov/cryptoreport.

146. ANTI-MONEY LAUNDERING AND COUNTERING THE FINANCING OF TERRORISM NATIONAL PRIORITIES, FINCEN (June 30, 2021), available at www.fincen.gov/sites/default/files/shared/AML_CFT%20Priorities%20(June%2030%2C%202021).pdf.

147. *See*, e.g., Katja Muñoz, *Metaverse: Personal Security, Regulation, New Global Order*, LINKEDIN (Nov. 4, 2021), https://www.linkedin.com/pulse/metaverse-personal-security-regulation-new-global-order-mu%C3%B1oz/; Mary Hui, *China is Eyeing the Metaverse as the Next Internet Battleground*, QUARTZ (Nov. 17, 2021), https://qz.com/2089316/china-sees-the-metaverse-as-the-next-internet-battleground/.

148. Hui, *supra* note 147 ("The metaverse will definitely become a wind vane of global technology development in the next decade, and will also become a new high ground of competition in the digital economy of all countries," said Wu, according to a summary of the event by CMCA.").

149. Brian Liu & Raquel Leslie, *As Facebook Joins the Race to the Metaverse, Chinese Tech Companies Face Hurdles*, LAWFARE (Nov. 12, 2021), https://www.lawfareblog.com/facebook-joins-race-metaverse-chinese-tech-companies-face-hurdles

> China's dominance in the artificial intelligence (AI) sphere may have given the country a sizable head start over competitors in the race to build the next frontier

regards data as a factor of production, and has erected a new legal infrastructure to ensure sweeping control over tech firms' data."[150] China is concerned that the U.S. companies that dominate the current internet will maintain or even extend that dominance into the metaverse.[151]

The race for global dominance of the metaverse is just a small part of its potential disruption in the context of radicalization, human rights abuses, and military aggression. The metaverse relies, in part, on the use of avatars and other forms of synthetic media to replace live video during the user's online experience.[152] As a result, the metaverse expands two destabilizing forces. First, interactions with avatars increase the potential that the other party is not a real person and is instead either paid by the belligerent aggressor or an AI deployed by that aggressor. Second, the synthetic media nature of the metaverse further undermines a user's visual and contextual cues for content veracity.

The use of avatars expands the ability of bad actors to create false identities for use in interacting with potential victims.[153] Since interacting with avatars will increasingly be the norm, it will become easier than ever for bad actors, such as trolls, to run different characters that each present as legitimate individuals. "Examples of these accounts include trolls, bots and sockpuppets, and all of them can make it difficult to identify legitimate sources of political discourse."[154]

> of virtual human interaction. China's New Generation Artificial Intelligence Development Plan has spurred billions of dollars in research and development investments from ministries, provincial governments and private companies since its issuance in 2017. China also has unrivaled capacity and experience in another driving force of the metaverse: consumer device manufacturing. ... Furthermore, Chinese interactive mobile platforms have a head start in innovation. ... Tencent presides over a sprawling business empire encompassing everything from gaming to mobile payments and virtual offices. ... Other Chinese tech giants have followed in Tencent's footsteps as competition in the space heats up. Just one day after Facebook announced its corporate name change to Meta, Chinese search engine Baidu applied to trademark the name "metaapp," while Chinese gaming giant NetEase filed dozens of trademark applications related to the buzzword. E-commerce giant Alibaba also registered several trademarks, including "Ali Metaverse".

150. Hui, *supra* note 147.

151. *Id.* ("In an interview last week with China News Network, Zuo [Pengfei, a researcher at the state-affiliated Chinese Academy of Social Sciences], cautioned that the metaverse has "an inherent monopoly gene," and that care must be taken to "avoid the metaverse being monopolized by a few forces."").

152. *See* Adi Robertson & Jay Peters, What is the metaverse, and do I have to care?, The Verge (Oct. 4, 2021), https://www.theverge.com/22701104/metaverse-explained-fortnite-roblox-facebook-horizon ("Gathering your co-workers around a virtual table in a service like Spatial and Facebook Horizon, for instance, could feel more natural to some people than looking at a grid of Zoom thumbnails.").

153. *See* John Silva, *Spotting Social Media 'Bad Actors*,' NEWS LITERACY PROJECT (Feb. 12, 2019), https://newslit.org/educators/civic-blog/spotting-social-media-bad-actors/ ("In the world of misinformation, a "bad actor" is a type of social media account that spreads misinformation and often causes confrontation.").

154. *Id.*

Dr. Ignas Kalpokas of Vytautas Magnus University, Lithuania, describes the phenomenon as a function of "the media's generative capacity."[155] The generative capacity for virtual worlds "refers to the capacity to create synthetic likenesses, personalities, and entire environments solely by way of digital technologies."[156]

Moving past the simple use of deepfakes to create the illusion that a real person is involved in some political scandal or pornographic activity, virtual worlds expand the potential for synthetic influencers to shape public opinion on any political agenda.

> Recent developments in today's media also involve the creation of synthetic personalities, primarily as virtual influencers (VIs). Like their human counterparts, these are personalities geared for maximum audience impact. However, due to their synthetic nature, VIs provide an unprecedented degree of flexibility and targeting. Hence, it is typical for creators to provide VIs with 'a composite personality based on market research,' and then use machine learning-based social listening to adapt to target audiences as effectively as possible.[157]

Through the use of AI, the particular message and approach of the VI can be adjusted to the emotional and cognitive triggers most likely to influence the user. Social media campaigns use psychographics to target the public based on each audience member's emotional profile.[158] The visual image of each avatar can be customized to enhance its affinity for the target. The key words and emotionally impactful phrases can be tailored to fit the target's profile. And the backstory of

---

> What is a 'troll'? This describes a person who deliberately posts offensive, inflammatory, highly partisan content in order to provoke people. . . . What is a 'sockpuppet'? This type of impostor account involves the creation of a false online identity, often to influence opinion about a person or organization with the intention of making it seem like the account is not affiliated in any way with that person or organization. . . . What is a 'bot'? Bots are 'automated user accounts that interact with Twitter using an application programming interface (API).' Think of it as a computer program that is designed to post content automatically according to a set of guidelines, without human intervention.

155. Ignas Kalpokas, *Problematising Reality: The Promises and Perils of Synthetic Media*, 2021 SOC. SCI. 1, 2 (Nov. 9, 2020), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7649059/.

156. *Id.*

157. *Id.* at 5 (quoting Sam Bradley, *Can Virtual Influencers build real Connections with Audiences?*, THE DRUM (Jan. 24, 2020) https://www.thedrum.com/news/2020/01/24/can-virtual-influencers-build-real-connections-with-audiences.

158. *What Is Psychographics? Understanding The Tech That Threatens Elections*, CB INSIGHTS (May 6, 2020), https://www.cbinsights.com/research/what-is-psychographics/ ("Psychographics is the study of consumers based on their activities, interests, and opinions (AIOs). ... Psychographics seeks to understand the cognitive factors that drive consumer behaviors. This includes emotional responses and motivations; moral, ethical, and political values; and inherent attitudes, biases, and prejudices.").

the avatar can be adjusted to make the VI more influential over each target.[159] When these sophisticated psychographic techniques are filtered through highly malleable avatars in the metaverse, they have the potential to build extremely large social movements extremely quickly, with the potential to unleash significant impact before the independent press or governmental agencies have any opportunity to challenge the veracity of the campaign. When targeted at the public's fears and prejudices, a well-honed campaign could trigger waves of anti-immigrant sentiment or motivate a state to secede from its European neighbors.

The use of synthetic media and falsified avatars then leads to the second threat from the metaverse. Once all media is artificial, at least in part, the public will struggle worse than ever to distinguish fact from fiction. "Detecting fake news can be difficult, especially when legitimate news organizations produce satirical programs easily mistaken for news reports. . . . Fake news articles often lack sources. . . . Often, legitimate organizations provide links to source information—fake news doesn't."[160] Even fiction is persuasive, particularly if it is immersive.

> Philosophers have long concerned themselves with what they call 'the paradox of fiction'—why would we find imagined stories emotionally arousing at all? The answer is that most of our mind does not even realize that fiction is fiction, so we react to it almost as though it were real.[161]

Taken together, the power of AI, synthetic media, and the synthetic metaverse in which to experience the deepfakes and other forms of deception has the potential to unleash political havoc if a dedicated belligerent nation or disruptive non-state actor invested the time and resources to do so. While AI and the metaverse are both beneficial technologies with clear, positive use cases, the potential for misuse cannot be ignored.

IV. CURRENT LEGAL CONSEQUENCES

At the moment, however, the misuse of AI, promulgation of synthetic media, and manipulation of social media and the virtual world have little, if any, legal consequences. An example of an earlier effort to interfere with the U.S. election in 2016 using Russia-backed trolls and a strategic campaign of disinformation has resulted in little criminal or international accountability.[162] In 2018, the Office of

---

159. *Id.* ("With its emphasis on the individual's personality traits, psychographic marketing reinforces the connection between product and personalization. People see themselves as themselves first — before they see themselves as members of an impersonal demographic group.").

160. Meagan Gillmore, *Fake News: Distinguishing Fact from Fiction*, TEACH MAG. (Mar./Apr. 2017), https://teachmag.com/archives/9860. ("People aren't directly quoted; source material for statistics may not be provided.").

161. Jim Davies, *Most of the Mind Can't Tell Fact from Fiction*, NAUTILUS (Sept. 11, 2019), https://nautil.us/blog/most-of-the-mind-cant-tell-fact-from-fiction.

162. *See* Ivan Nechepurenko and Michael Schwirtz, What We Know About Russians Sanctioned by the United States, N.Y. TIMES (Mar. 15, 2018), https://www.nytimes.com/2018/02/17/world/europe/russians-indicted-mueller.html.

Foreign Assets Control within the Department of Treasury sanctioned thirteen individuals.[163] These sanctions and an accompanying federal criminal indictment[164] were for serious international misconduct.

> Today's [indictments and sanctions] counter[] Russia's continuing destabilizing activities, ranging from interference in the 2016 U.S. election to conducting destructive cyber-attacks, including the NotPetya attack, a cyber-attack attributed to the Russian military on February 15, 2018 in statements released by the White House and the British Government. This cyber-attack was the most destructive and costly cyber-attack in history. Since at least March 2016, Russian government cyber actors have also targeted U.S. government entities and multiple U.S. critical infrastructure sectors, including the energy, nuclear, commercial facilities, water, aviation, and critical manufacturing sectors.[165]

Yet, despite the serious interference by Russia, the charges were later dropped against the two companies involved in the effort.[166] "Prosecutors said they concluded that a trial, against a corporate defendant with no presence in the United States and no prospect of meaningful punishment even if convicted, would likely expose sensitive law enforcement tools and techniques, 'potentially undermining

---

163. *Treasury Sanctions Russian Cyber Actors for Interference with the 2016 U.S. Elections and Malicious Cyber-Attacks*, U.S. DEPT. OF TREASURY PRESS RELEASE (Mar. 15, 2018), https://home.treasury.gov/news/press-releases/sm0312 (applying the "Countering America's Adversaries Through Sanctions Act (CAATSA) as well as Executive Order (E.O.) 13694, 'Blocking the Property of Certain Persons Engaging in Significant Malicious Cyber-Enabled Activities,' as amended, and codified pursuant to CAATSA.").

164. *See* U.S. v. Concord Management & Consulting, LLC, 317 F. Supp. 3d 598, 605–06 (D.D.C. 2018).

> According to the indictment, Concord or its co-conspirators "interfere[d] with the U.S. political system" by, among other things, "posing as U.S. persons and creating false U.S. personas," "operat[ing] social media pages and groups" that "falsely claimed to be controlled by U.S. activists," "us[ing] the stolen identities of real U.S. persons to post" on social media, "travel[ing] to the United States under false pretenses for the purpose of collecting intelligence," "procur[ing] and us[ing] computer infrastructure ... to hide the Russian origin of their activities and to avoid detection by U.S. regulators and law enforcement," "buying political advertisements on social media in the names of U.S. persons and entities," and "solicit[ing] and compensat[ing] real U.S persons" while "posing as U.S. grassroots entities and U.S. persons."

(internal references omitted).

165. *Treasury Sanctions Russian Cyber Actors for Interference with the 2016 U.S. Elections and Malicious Cyber-Attacks*, *supra* note 163.

166. Michael Balsamo and Eric Tucker, Feds Dropping Case For 2 Russian Companies in Troll Probe, FED. TIMES (Mar. 17, 2020), https://www.federaltimes.com/federal-oversight/doj-fbi/2020/03/17/feds-dropping-case-for-2-russian-companies-in-troll-probe/ ("The case was one of the signature indictments from Mueller's two-year Russia investigation. Together with a separate case against Russian government hackers accused of breaking into Democratic email accounts, it revealed a sweeping Russian effort to influence, or interfere in, the race between Republican Donald Trump and Democrat Hillary Clinton.").

their effectiveness.'"[167] In essence, there is little consequence for the conduct of these adversaries while they are protected by their home states.

While tangential to the disinformation campaign, Russia's NotPetya cyberattack "was the most destructive and costly cyber-attack in history. The attack resulted in billions of dollars in damage across Europe, Asia, and the United States, and significantly disrupted global shipping, trade, and the production of medicines."[168] The same sanctions covered the NotPetya attack as those imposed for the interference with the 2016 elections, meaning the same set of indictments and seizures that never had actual assets or individuals on which to impose the sanctions. Put another way, the worst cyberattack in history and the most significant effort to meddle in U.S. elections together combined for no penalties of any kind.

The potential for disruption will continue to grow as the metaverse expands in popularity and AI increases its influence. Without meaningful consequences to stop bad actors, the threats to disruption and the opportunities for attack are also likely to increase significantly.

A. The Law of War and International Law

The implementation of the global prohibition against genocide, crimes against humanity, war crimes, and the crime of aggression take place through the International Criminal Court (ICC), which was established by the Rome Statute of the ICC (Rome Statute) that entered into force on July 1, 2002.[169] By the terms of the treaty, however, the ICC only has jurisdiction over natural persons and excludes jurisdictions over corporations as well as governments, political parties, rebel movements, or other enterprises.[170] While the ICC could potentially reach the actions of individuals within corporations that participate in furthering genocide, the practical limitation that the court has no jurisdiction over the enterprise being directed by the person makes such liability unlikely. In addition, the ICC only has jurisdiction over those countries that have entered into the Rome Statute.[171] The United States has not.[172]

---

167. *Id.*

168. *Id.*

169. Rome Statute of the International Criminal Court, U.N. GAOR, U.N. Doc. A/CONF.183/9 (1998) (entered into force July 1, 2002) [hereinafter ROME STATUTE]; *see also* Göran Sluiter, *The Surrender of War Criminals to the International Criminal Court*, 25 LOY. L.A. INTL. & COMP. L. REV. 605 (2003) ("The ICC has jurisdiction over war crimes, crimes against humanity, and genocide.").

170. *See* ROME STATUTE Art. 25(1) ("The Court shall have jurisdiction over natural persons pursuant to this Statute."); *How the ICC Works*, AM. BAR ASS'N, https://how-the-icc-works.aba-icc.org (last visited Jan. 17, 2022) ("The ICC can only investigate and prosecute 'natural persons' who are over the age of 18. The ICC cannot investigate or prosecute governments, corporations, political parties, or rebel movements, but may investigate individuals who are members of groups.").

171. *See id.*

172. *See* Michael Scharf, *The ICC's Jurisdiction Over the Nationals of Non-Party States: A Critique of the U.S. Position*, 64 J. L. CONTEMP. PROB. 67 (2001), https://scholarlycommons.law.case.edu/faculty_publications/257

The scope of international customary law of non-intervention covers a very large grey area. The International Court of Justice has explained that intervention becomes coercive through the use of force or through "subversive or terrorist armed activities within another State."[173]

The U.N. Charter prohibits the "threat or use of force against the territorial integrity or political independence of any state, or in any other manner inconsistent with the Purposes of the United Nations."[174] Propaganda is not specifically addressed. However, "[t]his prohibition is complemented by a customary international law norm of nonintervention, which prohibits states from interfering in the internal affairs of other states."[175] Deceitful, destabilizing propaganda constitutes a direct interference, and as such violates the customary international law norm of non-intervention.[176]

The question then is whether media and communicative tools, when used fraudulently, can meet this standard for subversive action that violates the obligation of non-intervention.

> Whether a broadcast contravenes the non-intervention principle depends on all the circumstances. If it is deliberately false and intended to produce dissent or encourage insurgents, the non-intervention principle is likely to be breached. If factual and neutral, it is doubtful that the broadcast will constitute intervention, regardless of the effect it may in fact have.[177]

The growing importance of cyberattacks and cyberwarfare have led to the development of an academic precis on the law of cyberwar, The *Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations* (the Tallin Manual).[178] The Tallinn Manual notes that propaganda is not prohibited, and therefore is not itself an act of armed conflict.[179] At the same time, however, the

---

173. Military and Paramilitary Activities in and against Nicaragua (Nicaragua v. United States of America). Merits, Judgment. I.C.J. Reports 1986, p. 14. para 205.

> Intervention is wrongful when it uses methods of coercion in regard to such choices, which must remain free ones. The element of coercion, which defines, and indeed forms the very essence of, prohibited intervention, is particularly obvious in the case of an intervention which uses force, either in the direct form of military action, or in the indirect form of support for subversive or terrorist armed activities within another State.

174. U.N. Charter art. 2, ¶ 4.

175. Oona A. Hathaway, *et al.*, *The Law of Cyber-Attack*, 100 CAL. L. REV. 817, 842 (2012) (citing G.A. Res. 37/10, U.N. Doc. A/RES/37/10 (Nov. 15, 1982); G.A. Res. 25/2625, U.N. Doc. A/RES/25/2625 (Oct. 24, 1970).

176. *See* Björnstjern Baade, *Fake News and International Law*, 29 EUROPEAN J. OF INT'L L. 1357, 1363 (2018) ("Fake news is widely considered a substantial security threat, in particular, if it is state-sponsored. ... [Actions], such as incitement to revolutionary change, pass the threshold.").

177. Maziar Jamnejad & Michael Wood, *The Principle of Non-intervention*, 22 LEIDEN J. INT'L L. 345 (2009).

178. *See* MICHAEL N. SCHMITT (ED.), TALLINN MANUAL 2.0 ON THE INTERNATIONAL LAW APPLICABLE TO CYBER OPERATIONS (2017).

179. *Id.* at Rule 11 (9.)(h) ("international law does not prohibit propaganda, psychological operations, espionage, or mere economic pressure *per se*.").

promulgation of such propaganda can transform those who carry it out into military targets:

> The issue of whether the use of electronic or other media to spread propaganda qualifies as direct participation in hostilities (and the associated question of whether the objects used qualify as military objectives) is unsettled. The majority of the International Group of Experts took the position that broadcasts used to incite war crimes, genocide, or crimes against humanity render a journalist a direct participant and make the equipment used military objectives liable to attack, including by cyber means. A minority disagreed. The majority of the International Group of Experts also took the position that spreading propaganda does not pursue constitute direct participation in hostilities, while the minority suggested that the use of networks or computers to spread propaganda might convert journalistic equipment into a military objective for the purpose of cyber attacks. In any case, these issues are highly fact contingent.[180]

As recognized by the Tallinn Manual, neither the use of propaganda[181] nor the use of cyberattacks[182] is new in the field of international engagement or the modern era of the digital cold war.[183] Moreover, international law is largely toothless. The United Nations charter failed to address the potency of "pernicious propaganda" despite its critical role in both the first and second world wars.[184] Propaganda is

---

180. *Id.* at Rule 79 (9.) ("Civilian journalists engaged in dangerous professional missions in areas of armed conflict are civilians and shall be respected as such, in particular with regard to cyber attacks, as long as they are not taking a direct part in hostilities.").

181. *See e.g.*, Sarabeth A. Smith, *What's Old Is New Again: Terrorism and the Growing Need to Revisit the Prohibition on Propaganda*, 37 SYRACUSE J. INTL. L. & COM. 299, 302 (2010) ("Propaganda in its most neutral and simple sense is the persuasive dissemination of particular ideas or 'material disseminated by the advocates or opponents of a doctrine or cause.' However, under a modern understanding, to identify a message as propaganda is to 'suggest something negative and dishonest.'") (*quoting* dictionary.com and GARTH JOWETT & VICTORIA O'DONNELL, PROPAGANDA AND PERSUASION 2 (Sage Publications, 2006)).

182. *See* Daniel Garrie & Shane R. Reeves, *An Unsatisfactory State of the Law: The Limited Options for a Corporation Dealing with Cyber Hostilities by State Actors*, 37 CARDOZO L. REV. 1827 (2016) (discussing corporate liability for failing to stop nation-state cyberattacks such as the one perpetrated by North Korea on Sony in response to the release of its 2014 film, *The Interview*.).

183. *See* Oona A. Hathaway, *et al.*, *The Law of Cyber-Attack*, 100 CAL. L. REV. 817 (2012) ("The customary international law of countermeasures governs how states may respond to international law violations that do not rise to the level of an armed attack justifying self-defense—including, implicitly, cyber-attacks."); Eric Engle, *A New Cold War? Cold Peace. Russia, Ukraine, and NATO*, 59 ST. LOUIS U. L.J. 97, 112 (2014) ("The Cold War was characterized by constant conflict, overt and covert. Arms control was a perennial political issue of the Cold War to prevent or limit the arms race, and arms control remains a key issue with respect to Russia today.").

184. MICHAEL KEARNEY, THE PROHIBITION OF PROPAGANDA FOR WAR IN INTERNATIONAL LAW 55 (Oxford Univ. Press 2007) (citing Whitton, *supra* note 88 at 899); *see also* Natalie Maier, *Customary International Law As A Check on Press Freedom's Strongmen*, 47 SYRACUSE J. INTL. L. & COM. 305, 320 (2020) ("Adolph Hitler and his Nazi regime used some of the most powerful political rhetoric in history, despite its horrific consequences. . . . The use of such propaganda was crucial in 'defining the enemy,' and establishing the press as a threat to the security of the state.").

not outlawed under the U.N. Charter but is instead derived from the explicit ban on war.[185] "While propaganda was first prosecuted as an international crime during the Nuremburg Trials, it was not officially prohibited by international law until the adoption of Article 20 of the International Covenant on Civil and Political Rights (ICCPR) in 1966."[186] The ICCPR, together with the International Covenant on Economic, Social and Cultural Rights (ICESCR) serve to codify the Universal Declaration of Human Rights.[187]

ICCPR Article 20 provides simply: "1. Any propaganda for war shall be prohibited by law. 2. Any advocacy of national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence shall be prohibited by law."[188] Because any prohibition on propaganda implicates a restriction on speech, ICCPR Article 20 is preceded by Article 19, providing that "Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice."[189] The juxtaposition of the two articles creates a legal obligation to distinguish propaganda from legally protected free expression.

The enforcement of Article 20(1) generally has been enacted through domestic criminal laws.[190] "A large number of States parties claim to have satisfied the obligation of Article 20(1) by committing themselves to the object of peace, either through their constitutions, domestic legislation, or international pledges, without

---

185. KEARNEY, *supra* note 184 at 56.

> The contemporary prohibition of propaganda for war set forth in the International Covenant on Civil and Political Rights is derived from these fundamental principles of international law, namely, that war is outlawed and that the principle of freedom of expression cannot be abused in order to violate the rights and freedoms of others.

186. Sarabeth A. Smith, *supra* note 181 at 300 (citing International Covenant on Civil and Political Rights, Dec. 16 1966, 999 U.N.T.S. 171 [hereinafter ICCPR]).

187. *See* ICCPR. *See also* International Covenant on Economic, Social and Cultural Rights [hereinafter ICESCR].

188. ICCPR, Art. 20.

189. *Id.* at Art. 19. Article 19 has three subdivisions:

> 1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
> 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
> (a) For respect of the rights or reputations of others;
> (b) For the protection of national security or of public order (ordre public), or of public health or morals.

190. KEARNEY, *supra* note 184 at 134. ("A review of provisions of national law by which States parties have given effect to the prohibition demonstrates a strong inclination towards prohibiting propaganda for war through criminal rather than civil legislation.").

having enacted specific legislation concerning propaganda for war."[191] The prohibition against propaganda for war or incitement to discrimination is not directly prohibited by the Rome Statute of the International Criminal Court.[192] Although the U.S. has adhered to the ICCPR treaty,[193] it has done so with a package of reservations that makes it largely inapplicable to domestic law.[194] The U.S. specifically limited the impact of Article 20 to assure its compliance with the First Amendment by providing "[t]hat Article 20 does not authorize or require legislation or other action by the United States that would restrict the right of free speech and association protected by the Constitution and laws of the United States.[195]

Another source of international law can be found in the Convention on the Prevention and Punishment of the Crime of Genocide (Genocide Convention), which was adopted by the U.N. General Assembly in 1948.[196] In Article III(c) of the Genocide Convention, it prohibits "[d]irect and public incitement to commit genocide."[197] The Genocide Convention was later used as the template for the Statute of the International Criminal Tribunal for Rwanda and used for convictions

191. *Id.* at 139.

192. *Id.* at 191 ("More than sixty years later, the question of whether direct and public incitement to aggression constitutes a criminal act in contravention of international law has been tabled as part of the drafting of the crime of aggression for inclusion in the Rome Statute of the International Criminal Court."); *see also* Oona A. Hathaway, Paul K. Strauch, Beatrice A. Walton, & Zoe A. Y. Weinberg, *What Is A War Crime?*, 44 YALE J. INTL. L. 53, 98 (2019) ("A number of States in Europe explicitly tether parts of their domestic criminal codes concerning war crimes to international law. While some States include only the crimes outlined in Article 8 of the Rome Statute, the domestic statutes of many States include a much broader set of crimes as prosecutable 'war crimes.'").

193. U.S. RESERVATIONS, DECLARATIONS, AND UNDERSTANDINGS, INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS, 138 Cong. Rec. S4781-01.

194. *See* Louis Henkin, *U.S. Ratification of Human Rights Conventions: The Ghost of Senator Bricker*, 89 AM. J. INTL. L. 341, 341 (1995) ("the United States has attached to each of its ratifications a 'package' of reservations, understandings and declarations (RUDs), which has evoked criticism abroad and dismayed supporters of ratification in the United States. As a result of those qualifications of its adherence, U.S. ratification has been described as specious, meretricious, hypocritical.").

195. U.S. RESERVATIONS, DECLARATIONS, AND UNDERSTANDINGS, INTERNATIONAL COVENANT ON CIVIL AND POLITICAL RIGHTS, *supra* note 193; *see also* Paul J. Dombeck, *Imperative to Conscience: The Impact of Communications Media on the Practice of Genocide and Other War Crimes*, 1 CHI.-KENT J. INTL. & COMP. L. 1 (2001) ("In some form, freedom of expression is protected, or at least addressed in the constitutions of most nations. . . . In addition, '[t]he right to free speech stands as a general norm of customary international law. . . .'") (quoting DAVID JONES, HUMAN RIGHTS: GROUP DEFAMATION, FREEDOM OF EXPRESSION AND THE LAW OF NATIONS 37 (1998).

196. Convention on the Prevention and Punishment of the Crime of Genocide, Dec. 9, 1948, S. Exec. Doc. O, 81-1 (1949), 78 U.N.T.S. 277 [hereinafter Genocide Convention]; *see also* Gregory S. Gordon, *The Propaganda Prosecutions at Nuremberg: The Origin of Atrocity Speech Law and the Touchstone for Normative Evolution*, 39 LOY. L.A. INTL. & COMP. L. REV. 209, 239 (2017).

197. Genocide Convention, Art. III(c).

in those proceedings.[198] The approach was also used in trials for crimes against humanity in the former Yugoslavia.[199]

The manipulation of AI and synthetic media are not, of course, limited to propaganda. Through cyberattacks and other forms of unauthorized computer intrusions, a belligerent actor could retrain an AI to misunderstand critical information. Synthetic media could be used to masquerade as foreign leaders, civilians, or others, and a wide range of other misuses could be produced—some of which could result in direct casualties. Where AI and synthetic media are used in this manner, the Tallinn Manual would treat these as direct cyber attacks.[200]

Despite this limited set of examples, the history of international action against propaganda leading to war crimes, genocide, and crimes against humanity is very limited. The law has not proven effective against terrorist organizations, states committing human rights abuses against their citizens, or most of the instigators behind human rights abuses.[201] China, Russia (including the former Soviet Union), and other nations have not been held accountable for their use of propaganda or their systemic belligerent cyber operations. Given this history, existing international law is unlikely to provide much protection.

Although the international law has not yet proven effective to discourage corporate complicity with international atrocities, work continues to do so. The UN continues efforts to improve the customary international law through the General Principles on Business and Human Rights.[202] These General Principles provide a framework for expansion of international and domestic laws in this regard.

> (a) States' existing obligations to respect, protect and fulfil human rights and fundamental freedoms;
>
> (b) The role of business enterprises as specialized organs of society performing specialized functions, required to comply with all applicable laws and to respect human rights;
>
> (c) The need for rights and obligations to be matched to appropriate and effective remedies when breached.

---

198. *See* Prosecutor v. Akayesu, Case No. ICTR-96-4-T, Judgement, ¶ 550 (Sept. 2, 1998); Gordon, *supra* note 196 at 239-240.

199. *See* Prosecutor v. Kordić & Čerkez, Case No. IT-95-14/2-T, Judgment, ¶ 209 (Int'l Crim. Trib. for the Former Yugoslavia Feb. 26, 2001).

200. *See* TALLINN MANUAL at 91 ("The law of armed conflict applies to the targeting of any person or object during armed conflict irrespective of the means or methods of warfare employed. Consequently, the basic principles such as distinction and the prohibition of unnecessary suffering will apply to cyber operations just as they do to other means and methods of warfare.").

201. There is a much broader body of enforcement regarding parties involved in committing genocide and other atrocities, which is beyond the scope of this article. *See generally* Melissa Nobles, *The Prosecution of Human Rights Violations*, 13 ANNU. REV. POLIT. SCI. 165 (2010), https://www.annualreviews.org/doi/pdf/10.1146/annurev.polisci.040108.110013 (providing a survey of prosecutions and truth commissions involving outgoing authoritarian regimes).

202. *See* Ruggie, *supra* note 32.

> These Guiding Principles apply to all States and to all business enterprises, both transnational and others, regardless of their size, sector, location, ownership and structure.[203]

The General Principles represent a positive step in helping focus on the need for states to police their enterprises. The first foundational principle makes this clear. "States must protect against human rights abuse within their territory and/or jurisdiction by third parties, including business enterprises. This requires taking appropriate steps to prevent, investigate, punish and redress such abuse through effective policies, legislation, regulations and adjudication."[204] The General Principles provide a very useful approach to review U.S. domestic law, and the General Principles offer advice on how to improve adherence to these approaches.

B. Domestic Law in the U.S.

There is a wide range of domestic laws that cover the steps involving acts of murder, genocide, terrorism, and various war crimes. The purpose of this Article, however, is to focus only on the business organizations that facilitate these atrocities by assisting in their promotion of belligerent cyber operations, deceptive propaganda, and human rights abuses. The complex domestic law, customary international law, treaties, and international procedures for holding the perpetrators of war crimes and terrorism accountable are outside the scope of this review.[205] Since the text of ICCPR Article 20 is not self-executing positive law, the adherence to ICCPR does little to add to the U.S. legal basis for prohibiting falsified propaganda leading to genocide and human rights abuses.

Instead of looking to international law, there is a compelling case to look to U.S. law for at least a partial solution. No country has had a greater influence on the development of the internet and the potential impact of these technologies.[206]

---

203. *Id.* at 6.

204. *Id.* at 8.

205. *See generally* Alan F. Williams, *Overcoming the Unfortunate Legacy of Haditha, the Stryker Brigade "Kill Team," and Pantano: Establishing More Effective War Crimes Accountability by the United States*, 101 KY. L.J. 337, 344 (2013) ("'Grave breaches' [of the Geneva Convention] include willful killing, torture, or inhumane treatment, biological experiments, willfully causing great suffering or serious bodily injury or health, taking of hostages, unjustified and extensive destruction of property, compelling a prisoner of war (POW) to serve in the armed forces of his enemy, and willfully depriving a POW of his rights to a fair and regular trial."); *see also* Hiromi Sato, *The Separate Crime of Conspiracy and Core Crimes in International Criminal Law*, 32 CONN. J. INTL. L. 73, 98 (2016) ("Conspiracy as a separate crime has also been stipulated in multinational conventions on the regulation of certain types of crimes against humanity."); *see also* Oona A. Hathaway, Paul K. Strauch, Beatrice A. Walton, & Zoe A. Y. Weinberg, *supra* note 192 at 55.

206. *See e.g.*, *A Short History of the Internet*, SCIENCE+MEDIA MUSEUM (Dec. 3, 2020), https://www.scienceandmediamuseum.org.uk/objects-and-stories/short-history-internet; *see also* Kristin Delaney, *World Wide Web: Using Internet Governance Structures to Address Intellectual Property and International Development*, 32 BROOK. J. INTL. L. 603, 606 (2007) (discussing the "global sharing of knowledge," and noting that "much of the protected material is generated and owned in the United States"); *see also* Travis D. Shahan, *The World Summit on the Information*

Although China may rival the U.S. in the future development of AI, metaverse platforms, and synthetic media, the U.S. industry is likely to remain the leader in its field and establish international norms for these technologies.[207] In addition to enforcing anti-propaganda laws as a moral imperative to reduce genocide and human rights abuses, there is also a geopolitical need to show U.S. moral leadership in the growth of new technologies. If the U.S. operates in a fundamentally amoral manner, then the totalitarian regimes of China and Russia can assert successfully that there is no difference between the democratic regimes and non-democratic regimes.[208]

In most cases, the law focuses on the ability to criminally prosecute individuals who participate in criminal misconduct since domestic law does little directly against foreign states. The prosecution against Russia, for example, has instead been carried out by identifying individuals and organizational agents who are subject to criminal prosecution, seizure orders, and civil liability.[209] In the recent

---

Society and the Future of Internet Governance, 10 COMPUTER L. REV. & TECH. J. 325, 334 (2006) ("Since October 1998, the United States has asserted 'policy authority' over any changes to the root zone file.103 Control over the root zone file means control over the entire DNS, which can translate into significant influence on the Internet as a whole.").

207. *See* Guy Faulconbridge, *China has won AI battle with U.S., Pentagon's Ex-Software Chief Says*, REUTERS (Oct. 11, 2021), https://www.reuters.com/technology/united-states-has-lost-ai-battle-china-pentagons-ex-software-chief-says-2021-10-11/ ("China, the world's second largest economy, is likely to dominate many of the key emerging technologies, particularly artificial intelligence, synthetic biology and genetics within a decade or so, according to Western intelligence assessments."); *see also* Martin Wolf, *China Battles the US in the Artificial Intelligence Arms Race*, FIN. TIMES (Apr. 16, 2019), https://www.ft.com/content/2f295a9e-5f96-11e9-b285-3acd5d43599e ("China's hypercompetitive and entrepreneurial economy lives by Facebook founder Mark Zuckerberg's notorious motto: 'move fast and break things.' Mr Lee describes a world of cut-throat business activity and remorseless imitation, which has already allowed Chinese businesses to defeat leading western rivals in their home market."); *see also* Catherine D. Henry, *Is China building the Metaverse?*, TECHCRUNCH (Nov. 1, 2021), https://techcrunch.com/2021/11/01/is-china-building-the-metaverse/ ("both the U.S. and China are trying to build and lay claim to the metaverse. . . . Competence and dominance across these four criteria is what may give China an insurmountable head start over the U.S. in the race to build the future of the virtualized human experience.").

208. *See* Barbara Lippert & Volker Perthes, *Strategic Rivalry between United States and China*, STIFTUNG WISSENSCHAFT UND POLITIK (June 4, 2020), https://www.swp-berlin.org/en/publication/strategic-rivalry-between-united-states-and-china ("the systemic conflict will loom increasingly large on the American side, sometimes interpreted as a clash between 'liberal democracy' and what is occasionally referred to as 'digital authoritarianism.' Highlighting the ideological conflict might be employed to mobilise sustained domestic support for a power clash with China. . ."); *see also* Eugene Rumer & Richard Sokolsky, *Thirty Years of U.S. Policy Toward Russia: Can the Vicious Circle Be Broken?*, CARNEGIE ENDOWMENT FOR INT'L PEACE (June 20, 2019), https://carnegieendowment.org/2019/06/20/thirty-years-of-u.s.-policy-toward-russia-can-vicious-circle-be-broken-pub-79323 ("Russian leaders see their country as a great power in charge of its own destiny. . . . [T]hey reject democracy promotion as a cover for U.S.-sponsored regime change; they . . . will resist perceived U.S. intrusions; and they rely on anti-Americanism to legitimize their unpopular policies with domestic audiences.").

209. *See* U.S. v. Andrienko 2:20-cr-00316 (W. D. Penn 2020), available at https://www.pacermonitor.com/public/case/37098649/USA_v_ANDRIENKO_et_al.

indictment against Russia's premiere cyberattack unit, the Main Intelligence Directorate of the General Staff of the Armed Forces (GRU),[210] the actual charges had little to do with cyber warfare. Instead, the charges included wire fraud, damage to computers, identity theft, criminal conspiracy, and aiding and abetting.[211]

Congress has been frustrated that little can be done against the nations responsible for attacks on America and its citizens. Reflecting this frustration, in 2016, Congress overrode a presidential veto to pass the Justice Against Sponsors of Terrorism Act (JASTA).[212] JASTA abrogates sovereign immunity to provide a terrorism victim the ability to bring tort claims against foreign states.[213]

> A foreign state shall not be immune from the jurisdiction of the courts of the United States in any case in which money damages are sought against a foreign state for physical injury to person or property or death occurring in the United States and caused by--
>
> (1) an act of international terrorism in the United States; and
>
> (2) a tortious act or acts of the foreign state, or of any official, employee, or agent of that foreign state while acting within the scope of his or her office, employment, or agency, regardless where the tortious act or acts of the foreign state occurred.[214]

The purpose of JASTA

> is to provide civil litigants with the broadest possible basis, consistent with the Constitution of the United States, to seek relief against persons, entities, and foreign countries, wherever acting and wherever they may be found, that have provided material support, directly or indirectly, to foreign

---

> GRU officers working for Military Unit 74455 ... knowingly and intentionally conspired ... to deploy destructive malware and take other disruptive actions ...to undermine, retaliate against, or otherwise destabilize: (1) Ukraine; (2) the country of Georgia; (3) France's elections; (4) efforts to hold Russia accountable for its use of a weapons-grade nerve agent on foreign soil; and (5) the 2018 Winter Olympics after a Russian government-sponsored doping effort led to Russian athletes being unable to participate under the Russian flag.

see also Countering America's Adversaries Through Sanctions Act (CAATSA), Pub. L. No. 115-44, 131 STAT. 886 (2017); *see also* Executive Order (E.O.) 13694, *supra* note 163; *see also* U.S. v. Concord Management & Consulting, LLC, 317 F. Supp. 3d 598, 605–06 (D.D.C. 2018).

210. U.S. v. Andrienko, *supra* note 209 at 1.

211. *Id.* (18 U.S.C. §§ 371, 1030(a)(2)(C), 1030(a)(5)(A), 3559(g)(l) (Conspiracy); 18 U.S.C. § 1349 (Conspiracy to Commit Wire Fraud); *see also* 18 U.S.C. § 1343 (Wire Fraud); *see also* 18 U.S.C. § 1030(a)(5)(A) and 1030(c)(4)(B) (Damage to Computers); *see also* 18 U.S.C. § 1028A (Aggravated Identity Theft); *see also* 18 U.S.C. § 2 (Aiding and Abetting).

212. Pub. L. No. 114-222, 130 Stat. 852 (2016) (codified at 28 U.S.C.A. § 1605B) [hereinafter JASTA].

213. 28 U.S.C.A. § 1605B (West) (Responsibility of foreign states for international terrorism against the United States.).

214. *Id.*

> organizations or persons that engage in terrorist activities against the United States.[215]

JASTA may prove useful for establishing direct liability if cyberattacks launched at the U.S. result in direct harm similar to the types of injuries, death, and property destruction typical of terrorist attacks. Yet, despite the broad scope of JASTA tort liability, the law does not adequately address the challenge posed by harmful propaganda, which often supports the physical acts of terrorism. The definition of terrorism requires that it include "violent acts" that would be criminal if carried out inside the U.S.[216] The law also excludes "any act of war."[217]

The scope of the law does cover "material support," which might include "the use of networks or computers to spread propaganda."[218] However, as noted earlier, only a minority of the International Group of Experts took the position that the equipment used to spread propaganda might be treated as a proper military objective in the Tallinn Manual.[219] It is even more tenuous in the context of terrorism. As a result, the state sponsors of propaganda are not likely to face legal responsibility under JASTA. Given the low risk of consequence and the high opportunity for disruption, the use of false propaganda is likely to increase in the metaverse as the metaverse gains popularity and becomes ever easier to propagate using synthetic media and AI.

Congress has supplemented criminal statutes with civil tort remedies, including treble damages and attorneys' fees under the Anti–Terrorism Act (ATA).[220] "Liability under the ATA has three elements: (1) unlawful action, i.e. an "act of international terrorism;" (2) the requisite mental state, and (3) causation."[221] Acts of international terrorism are broadly defined to reflect criminal activities intended to coerce a civilian population.[222] Congress has also enacted

---

215. JASTA, at § 2(b); *see also* In re Terrorist Attacks on September 11, 2001, 298 F. Supp. 3d 631, 642 (S.D.N.Y. 2018).

216. 18 U.S.C. § 2331 (2018) ("the term 'international terrorism' means activities that—(A) involve violent acts or acts dangerous to human life that are a violation of the criminal laws of the United States or of any State, or that would be a criminal violation if committed within the jurisdiction of the United States or of any State. . .").

217. 28 U.S.C.A. § 1605B. Act of war is defined in 18 U.S.C. § 2331 as "(A) declared war; (B) armed conflict, whether or not war has been declared, between two or more nations; or (C) armed conflict between military forces of any origin."

218. TALLINN MANUAL at Rule 71 (9).

219. *See id.*

220. 18 U.S.C. § 2333 (2021)

> Any national of the United States injured in his or her person, property, or business by reason of an act of international terrorism, or his or her estate, survivors, or heirs, may sue therefor in any appropriate district court of the United States and shall recover threefold the damages he or she sustains and the cost of the suit, including attorney's fees.

221. In re Chiquita Brands Int'l., Inc., 284 F. Supp. 3d 1284, 1304 (S.D. Fla. 2018) (citing Sokolow v. Palestine Liberation Org., 60 F.Supp.3d 509, 514 (S.D.N.Y. 2014)).

222. § 2331(1), as activities that:

federal statutes specifically targeting material support to terrorists (18 U.S.C. §2339A) who commit specified crimes[223] and prohibiting persons from knowingly providing material support or resources to "foreign terrorist organizations (18 U.S.C. §2339B)."[224] Since first passed in 1994, §2339A and §2339B have been expanded by congress numerous times to increase the jail terms, redefine "material support or resources," and incorporate "expert advice or assistance" to the list of prohibited support activities.[225]

Under both sections of the law, the term "material support or resources" is defined to be:

> any property, tangible or intangible, or service, including currency or monetary instruments or financial securities, financial services, lodging, training, expert advice or assistance, safehouses, false documentation or identification, communications equipment, facilities, weapons, lethal substances, explosives, personnel (1 or more individuals who may be or include oneself), and transportation, except medicine or religious materials.[226]

Section 2339A is a broad prohibition against providing terrorists material support, but it is limited in scope to specified crimes, such as biological or chemical weapons offenses; assassination, kidnapping, or assaulting the President, Vice-President, Members of Congress, the Supreme Court, or the Cabinet; transactions involving nuclear material; multinational acts of terrorism; bombing public places

---

(A) involve violent acts or acts dangerous to human life that are a violation of the criminal laws of the United States or of any State, or that would be a criminal violation if committed within the jurisdiction of the United States or of any State;
(B) appear to be intended—
(i) to intimidate or coerce a civilian population;
(ii) to influence the policy of a government by intimidation or coercion; or
(iii) to affect the conduct of a government by mass destruction, assassination, or kidnapping; and
(C) occur primarily outside the territorial jurisdiction of the United States, or transcend national boundaries in terms of the means by which they are accomplished, the persons they appear intended to intimidate or coerce, or the locale in which their perpetrators operate or seek asylum.

223. 18 U.S.C. § 2339A (2021) (Providing Material Support to Terrorists).

224. 18 U.S.C. § 2339B (2021) (Providing material support or resources to designated foreign terrorist organizations); *see also* Charles Doyle, *Terrorist Material Support: An Overview of 18 U.S.C. §2339A and §2339B*, CONG. RES. SVS. 7-5700 (Dec. 8, 2016), https://www.justice.gov/archives/jm/criminal-resource-manual-15-providing-material-support-terrorists-18-usc-2339a ("The two federal material support statutes have been at the heart of the Justice Department's terrorist prosecution efforts. One provision outlaws providing material support for the commission of certain designated offenses that might be committed by terrorists, 18 U.S.C. §2339A. The other outlaws providing material support to certain designated terrorist organizations, 18 U.S.C. §2339B.").

225. *See* Doyle, *supra* note 224, at 2.

226. 18 U.S.C. § 2339A(b). In addition, the terms "training" and "expert advice and assistance" are further defined: "(2) the term "training" means instruction or teaching designed to impart a specific skill, as opposed to general knowledge; and (3) the term "expert advice or assistance" means advice or assistance derived from scientific, technical or other specialized knowledge." *Id.*

or infrastructure facilities; or similar offenses.[227] "After the Oklahoma City bombing, Congress passed section 2339B, which bars material support of terrorist groups, such as Hamas."[228]

Both sections 2339A and 2339B are criminal statutes carrying significant jail terms, so courts interpreting these statutes impute a *mens rea* requirement that the material support be provided on a knowing basis by the accused.[229]

> [I]n contrast to § 2339B, which broadly criminalizes the provision of "material support" to formally designated foreign terrorist organizations, and requires knowledge about the organization's connection to terrorism, but not a specific intent to further its terrorist activities, Section 2339A "raises the scienter requirement" and criminalizes material support only where the

227. *See* Doyle, *supra* note 224, at 6 for a complete list of included offenses:

> 18 U.S.C. §2339A(a) ("Whoever provides material support … knowing or intending that [it is] to be used … in carrying out a violation of section 32 [destruction of aircraft and aircraft facilities], 37 [violence at international airports], 81 [arson within the special maritime and territorial jurisdiction of the United States, 175 [biological weapons offenses], 229 [chemical weapons offenses], 351 [assassination, kidnapping, or assaulting Members of Congress, the Supreme Court, or the Cabinet], 831 [transactions involving nuclear material], 844(m) [importing or exporting plastic explosives without a detection agent], 842(n) [possession of a plastic explosive without a detection agent], 844(f) [bombing federal property], 844(i) [bombing property used in, or affecting, interstate or foreign commerce], 930(c) [killing a person in the course of an attack on a federal facility with a firearm or dangerous weapon], 956 [conspiracy to kill, kidnap, maim, or injure individuals, or to damage property, in a foreign country], 1091 [genocide], 1114 [killing a federal officer, employee, or member of the armed forces], 1116 [killing internationally protected individuals], 1203 [hostage taking], 1361 [destruction of federal property], 1362 [destruction of communication lines, stations or systems], 1363 [destruction of property in the special maritime and territorial jurisdiction of the United States], 1366 [destruction of an energy facility], 1751 [assassination, kidnapping, or assaulting of the President, Vice President, or senior White House staff members], 1992 [terrorist on mass transit), 2155 [destruction of national defense material], 2156 [production of defective national defense material), 2280 [violence against maritime navigation], 2281 [violence against maritime fixed platforms], 2332 [killing or assaulting a United States national outside the United States], 2332a [use of weapons of mass destruction], 2332b [multinational acts of terrorism], 2332f [bombing public places or infrastructure facilities], 2340A [torture abroad], or 2442 [recruiting or using child soldiers] of this title; section 236 of the Atomic Energy Act of 1954 (42 U.S.C. 2284) [sabotage of nuclear facilities or fuel]; section 46502 [aircraft piracy] or 60123(b) [destruction of gas pipelines] of title 49 ... ").

228. Peter Margulies, *Defining, Punishing, and Membership in the Community of Nations: Material Support and Conspiracy Charges in Military Commissions*, 36 FORDHAM INTL. L.J. 1, 57–58 (2013).

229. *See* Morissette v. United States, 342 U.S. 246, 251 (1952) ("Crime, as a compound concept, generally constituted only from the concurrence of an evil-meaning mind with an evil-doing hand …"); U.S. v. Harcevic, 999 F.3d 1172, 1177 (8th Cir. 2021) ("Section 2339A makes it a crime to knowingly or intentionally supply 'material support or resources' in violation of one of a lengthy list of statutes. . . ").

> defendant acts with actual knowledge or intent that the support will be used to prepare for, or carry out, certain terrorism-related crimes. So, where § 2339A serves as the predicate ATA crime, an ATA plaintiff must prove that the defendant acted with the specific knowledge or intent that its support would be used in preparation for, or in carrying out, one of the enumerated terrorism-related crimes. On the other hand, it is not necessary for an ATA plaintiff to show the defendant's "specific intent to aid or encourage the particular attacks that injured plaintiffs."[230]

Both 2339A and 2339B have withstood constitutional challenges.[231] In addressing §2339B, the Supreme Court found the law constitutional while clarifying the *mens rea* requirement:

> Section 2339B(a)(1) prohibits "knowingly" providing material support. It then specifically describes the type of knowledge that is required: "To violate this paragraph, a person must have knowledge that the organization is a designated terrorist organization ..., that the organization has engaged or engages in terrorist activity ..., or that the organization has engaged or engages in terrorism ...." Congress plainly spoke to the necessary mental state for a violation of § 2339B, and it chose knowledge about the organization's connection to terrorism, not specific intent to further the organization's terrorist activities.[232]

Although these provisions have been upheld for direct legal action against individual terrorists, they have been less effective at holding accountable those who knowingly provide material support.[233]

---

230. In re Chiquita Brands Intl., Inc., 284 F. Supp. 3d 1284, 1309 (S.D. Fla. 2018) (quoting United States v. Awan, 459 F.Supp.2d 167 (E.D.N.Y. 2006)), aff'd, 384 Fed. Appx. 9 (2d Cir. 2010); *see* also Rothstein v. UBS AG, 708 F.3d 82 (2d Cir. 2013); *see also* Boim v. Holy Land Foundation for Relief and Development, 549 F.3d 685 (7th Cir.2008) (en banc); *see also* United States v. Al-Hussayen, 2004 U.S. Dist. LEXIS 29793, Case No. CR03-048-C-EJL, slip. op. (D. Id. 2004). *But see* United States v. Sattar 272 F.Supp.2d 348, 355 (S.D. N.Y. 2003).

231. Holder v. Humanitarian L. Project, 561 U.S. 1 (2010) (§ 2339B); *see also* U.S. v. Hassan, 742 F.3d 104, 129 (4th Cir. 2014) (citing § 2339A).

232. Holder v. Humanitarian L. Project, 561 U.S. 1, 16–17 (2010).

233. *See* Boim v. Holy Land Found. for Relief and Dev., 549 F.3d 685, 700 (7th Cir. 2008). Although there were many infirmities in the case against the Holy Land Foundation, the Seventh Circuit extend liability to the material supporter of terrorism:

> If the financier knew that the organization to which it was giving money engaged in terrorism, penalizing him would not violate the First Amendment. Otherwise someone who during World War II gave money to the government of Nazi Germany solely in order to support its anti-smoking campaign could not have been punished for supporting a foreign enemy.

See also Weiss v. Natl. Westminster Bank, PLC., 993 F.3d 144, 163 (2d Cir. 2021) ("In order for a plaintiff to prevail on an ATA claim against a defendant as a principal, the elements listed in § 2333(a) must be proven; an element is not proven unless the evidence comports with the ATA's definition of the element. . . "); *see also* Siegel v. HSBC North America Holdings, Inc., 933 F.3d

In the context of media companies, there is yet another barrier to finding liability even when a company knowingly provides the technological tools for propaganda. The Ninth Circuit has applied the Communications Decency Act (CDA) § 230[234] to such activities, creating a statutory bar to liability for materially aiding terrorist activities.[235] *Force v. Facebook* involved claims regarding the murders by Hamas of Yaakov Naftali Fraenkel, Chaya Zissel Braun, Richard Lakin, and Taylor Force, and attempted murder of Menachem Mendel Rivkin. These attacks all occurred against U.S. citizens while in Israel.[236] The Ninth Circuit barred claims under JASTA against Facebook pursuant to CDA § 230.

Gonzales v. Google involved claims that ISIS carried out a series of terrorist attacks, including murders in Paris, Istanbul, and San Bernardino. In these attacks, Nohemi Gonzalez, Nawras Alassaf, Sierra Clayborn, Tin Nguyen, and Nicholas Thalasinos lost their lives, and their families brought legal claims against the social media firms used by ISIS to promote its terrorist and propaganda campaign.[237]

At the heart of the complaint is the assertion that Google, Twitter, and Facebook are secondarily liable under 18 U.S.C. § 2333(d) for aiding and abetting an act of international terrorism and for conspiring with a perpetrator of an act of international terrorism. The families did not bring their claims together, but they were consolidated on appeal. Only Taamneh asserted direct liability under § 2333(a) for providing material support and resources to ISIS, and for concealing this support, in violation of 18 U.S.C. §§ 2339A, 2339B(a)(1), and 2339C(c).[238] None of these claims were successful.

In *Force v. Facebook*, the court reviewed Facebook's terms of service and prohibitions against terrorism. It noted that Facebook has a detailed policy prohibiting terrorist organizations from maintaining a presence on the site. "Facebook 'do[es] not allow symbols that represent any [terrorist] organizations or [terrorists] to be shared on [the] platform without context that condemns or neutrally discusses the content.'"[239] The Facebook Community Standards, as well as similar community standards provisions used by Google and Twitter, which

---

217, 224 & n.6 (2d Cir. 2019) (aiding-and-abetting liability requires a defendant to be "aware, based on public reports, that its banking customer was believed by some to have links to terrorist organizations."); *see also* Linde v. Arab Bank, PLC, 882 F.3d 314 (2d Cir. 2018); *see also* Margulies, *supra* note 228, at 58. ("Courts construing section 2339B have uniformly upheld Congress's view that [designated as foreign terrorist organizations], like state sponsors of terrorism, 'are so tainted by their criminal conduct that any contribution to such an organization facilitates that conduct.'") (quoting statutory findings).

234. See 47 U.S.C. § 230(c) (2021).

235. Gonzalez v. Google LLC, 2 F.4th 871, 880 (9th Cir. 2021) (rejecting liability under the Anti-Terrorism Act (ATA) for terrorist acts by ISIS, 18 U.S.C. § 2333); *see also* Force v. Facebook, Inc., 934 F.3d 53 (2d Cir. 2019) (In a claim involving terrorism by Hamas, the court rejected plaintiff's assertion that JASTA requires CDA § 230 to exclude material support for terrorist video and social media hosting).

236. Force, 934 F.3d at 57-58.

237. Gonzalez, 2 F.4th at 884.

238. *Id.*

239. Force, 934 F.3d at 60 (quoting Facebook's Community Standards).

were litigated in *Gonzales*, all condemn terrorism and prohibit the use of the platform for such services.

Given the broad usage of Facebook, Google (including both search and YouTube), and Twitter, it is highly unlikely that any of these companies could be found to have intentionally provided material support to commit terrorism. On the other hand, it is certainly possible that these companies were negligent in failing to remove content from their sites through the use of algorithms that promoted the terrorist messages to users of the sites, or when monetizing this content and thereby normalizing it for some members of the community. But JASTA requires knowing material participation. It may allow for liability for reckless disregard of the facts demonstrating the terrorist's organization's use of the corporate facilities to materially aid the terrorism, but that distinction has yet to be challenged in court.

The Ninth Circuit has consistently interpreted CDA § 230 in the broadest possible manner.[240] In both *Force* and *Gonzales*, the Ninth Circuit panels used § 230 to shield any liability for the content posted in support of terrorist activities. While the result may be correct for these companies, it would be wholly inconsistent if applied to a person or company that committed atrocities in Myanmar, Israel, Afghanistan, or Europe simply because the content was posted anonymously by terrorists outside the U.S. *Force* and *Gonzales* confuse companies that did not intentionally or recklessly and materially aid in terrorism with a legal analysis that says media platforms cannot materially aid in terrorism unless they create and post the content.[241] If CDA § 230 were appropriately used to "bar lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions," then it would fulfill its intended congressional purpose and still permit claims against those companies that intentionally—and perhaps recklessly—facilitate terrorism through the propaganda they permit on their platforms.

The frustration with the current litigation should not automatically equate to an assumption that any of these companies were guilty of providing material support to terrorists. According to an ad run by Meta to defend its efforts on content moderation, the company claims that "[w]e've invested more than $13 billion in teams and technology to stop bad actors and remove illicit content. Since July

240. Force, 934 F.3d at 80 (2d Cir. 2019) (Katzmann, Chief Judge, concurring) ("Illuminating Congress's original intent does, however, underscore the extent of § 230(c)(1)'s subsequent mission creep. Given how far both Facebook's algorithms and plaintiffs' terrorism claims swim from the shore of congressional purpose, caution is warranted before courts extend the CDA's reach any further."). For a discussion of CDA's unintended expansion, see Jon M. Garon, *Constitutional Limits on Administrative Agencies in Cyberspace*, 8 BELMONT L. REV. 499 (2021); *see also* Jon M. Garon, *Dysregulating the Media: Digital Redlining, Privacy Erosion, and the Unintentional Deregulation of American Media*, 72 ME. L. REV. 45 (2020).

241. *See* Force, 934 F.3d at 81 (2d Cir. 2019) (Katzmann, Chief Judge, concurring)

> Accordingly, our precedent does not grant publishers CDA immunity for the full range of activities in which they might engage. Rather, it "bars lawsuits seeking to hold a service provider liable for its exercise of a publisher's traditional editorial functions—such as deciding whether to publish, withdraw, postpone or alter content" provided by another for publication.

[2021], we've taken action on: 1.8 billion fake accounts, 26.6 million violent and graphic posts, [and] 9.8 million terrorism-related posts."[242] The scale of the current problem with hate speech, fake accounts, and disinformation is enormous. And this is the tip of the iceberg. Once the metaverse becomes the standard environment for business, commercial, communal, and educational interactions rather than just as a place to play, the potential exists for AI and synthetic media to trigger massive social upheaval. While this might cost billions of dollars to stop, the risk of harm could be vastly larger.[243]

V. MOVING FORWARD—CREATING A DUTY OF CARE FOR METAVERSE OPERATORS AND AI PUBLISHERS

As the preceding discussion suggests, the present state of anti-terrorism law is insufficient to address the potential storm of propaganda that may arise in some corners of the metaverse, triggered by the ease with which synthetic media and AI-fueled hate can be programmed and distributed. The efforts by the largest media companies are reported to cost billions to compete with the hate speech, trolling, disinformation, synthetic media, and falsified accounts used on those sites.

The potential for a decentralized Web3 will be undone if it becomes a breeding ground for false propaganda, state-sponsored cyberattacks, terrorism, and other forms of misconduct. When the Web 1.0 wild west opened, there was little use by terrorist organizations and cybercriminals. Instead, the perceived threats were from teenage hackers who intended mischief rather than civil unrest.[244] DAOs and decentralized services may be ill-equipped to address the constant onslaught of negative content. Or not. There is no particular reason that these services cannot operate without the harangue of negativity that plagues the largest media platforms. As governance for these platforms, however, the duty is to prevent the harm by planning for the threat of misuse rather than providing empty platitudes after the harm to civil society has been done.

Neither the ATA nor JASTA creates meaningful incentives to require U.S. businesses to operate proactively to stop synthetic media or AI-misuse from creating even wide-scale social misinformation campaigns, and when those

---

242. Facebook Ad, distributed via Axios Twin Cities (Jan. 4, 2022) (on file with author); *see also* David Pierce, *Facebook's Whistleblower Speaks*, SOURCE CODE (Oct. 4, 2021), https://www.protocol.com/newsletters/sourcecode/facebooks-whistleblower-speaks?rebelltitem=1#rebelltitem1 (quoting similar claims by Facebook while noting that Facebook's ads are in response to widespread accusations of failure to protect users and the public).

243. *See id.* (In a response to the testimony whistleblower, Frances Haugen, Nick Clegg, Facebook's vice president of policy and global affairs, wrote "'[t]he prevalence of hate speech on our platform is now down to about 0.05%.' . . . That's great! And it still means millions of hate-speech posts are flowing through the platform at any given time.").

244. *See* Jacob Savage, *Coming of Age with the Internet: Remembering Web 1.0*, AM. READER, https://theamericanreader.com/coming-of-age-with-the-internet-remembering-web-1-0/ (last visited Jan. 5, 2022) ("In the mid 90's, hackers ruled supreme. There were newspaper articles about hackers, movies about hackers, after-school specials about hackers. They were superhuman wizards who could bring down the U.S. government with just eight keystrokes.").

campaigns are used as propaganda to attack vulnerable minorities, wage civil war, or facilitate incursions into sovereign territory, then something must be done.

Ideally, congress should act to provide specific legislation by amending the ATA and JASTA. Sections 2339A and 2339B of JASTA do not create tort liability, so presently, plaintiffs must connect the violation of § 2339A or §2339B to treble damages under § 2333(a).[245] Since the goal of the proposal is to encourage platform providers to do more to thwart misconduct on their systems, the FTC can be granted additional authority by Congress to require affirmative steps be taken to stop the use of company resources in violation of § 2339A or §2339B. Through statute or regulation, the FTC could require annual reports by covered platforms for public information on the steps taken to meet the obligations to avoid providing material support for terrorism or to terrorist organizations under § 2339B.

The proposal is a modest step. It does not suggest expanding the scope of §2333 and adding tort liability for failing to stop terrorism. But it does have an influence on civil tort liability because a company that has an affirmative duty to provide annual reports is more likely to be knowingly aware of what it is doing and failing to do to stop terrorists from using its system. This approach is consistent with the General Principles on Business and Human Rights, which suggests adding "human rights due diligence" as part of domestic regulation.[246]

At the same time, Congress is updating JASTA to include an affirmative duty to stop the use of platforms in furtherance of terrorism or in support of terrorist organizations; it could perhaps expand or clarify the mens rea requirement for direct liability under §§ 3339A and 2339B to include only a scienter requirement so that enterprises that fail to act in reckless disregard of the facts can be liable for their nonaction under both criminal law and through civil administrative action.[247]

---

245. *See* Weiss, 993 F.3d at 160 ("Section § 2339B, while making the provision of material support or resources to an FTO a crime, does not itself provide a private right of action; the civil action is authorized by § 2333(a).").

246. Ruggie, *supra* note 32, at 11.

> States should take additional steps to protect against human rights abuses by business enterprises that are owned or controlled by the State, or that receive substantial support and services from State agencies such as export credit agencies and official investment insurance or guarantee agencies, including, where appropriate, by requiring human rights due diligence.

247. *See, e.g.,* Lorenzo v. Securities and Exch. Commn., 139 S. Ct. 1094 (2019) (applying scienter requirement); *see also* Ernst & Ernst v. Hochfelder, 425 U.S. 185, 212–214, (1976) (applying scienter requirement to Rule 10b–5(b)); *see also* Securities and Exch. Comm'n. v. Zenergy Int'l., Inc., 430 F. Supp. 3d 384, 394 (N.D. Ill. 2019) ("'[D]eliberate ignorance' satisfies the scienter requirement of the securities laws.") (citing SEC v. Jakubowski, 150 F.3d 675, 681–82 (7th Cir. 1998); *see* also In re Alphabet, Inc. Securities Litig., 1 F.4th 687, 699 (9th Cir. 2021) (quoting Janus Cap. Grp., Inc. v. First Derivative Traders, 564 U.S. 135, 142 (2011)).

> Under Section 10(b) and Rule 10b-5(b), 'the maker of a statement is the person or entity with ultimate authority over the statement, including its content and whether and how to communicate it.' Persons 'who do not 'make' statements (as Janus defined 'make'), but who disseminate false or misleading statements to potential investors with the intent to defraud, can be found to have violated the other parts of Rule 10b-5, subsections (a) and (c), as well as related provisions of the securities laws' including Section 10(b).

In the absence of congressional action, the FTC may take the position that it already has sufficient authority under Section 5 of the Federal Trade Commission Act (FTC Act) to determine that the failure to stop the deceptive propaganda and use by terrorist organizations constitutes an unfair and deceptive trade practice.[248] The FTC has successfully extended § 5 to failures to provide cybersecurity protection,[249] and failure to stop terrorist activities is certainly as egregious. Further, to the extent a company publicly states it follows its published content and user guidelines but fails to do so, it then acts in a manner that is deceptive to the users of the platform and the public.[250]

These steps will help address the expanded threat vector from terrorist organizations, but they do nothing to address the state actors themselves. Instead, Congress must enact a new provision under JASTA or elsewhere that creates the same obligation to defend against state actors and acts of war as the proposal requires for acts of terrorism.[251] As presently written, JASTA excludes acts of war and nation-states.[252] The proposed new regulations that require an affirmative obligation to stop misuse of platforms by terrorists and belligerent nations will need to expand beyond JASTA since JASTA is intended to be narrower.

Congress has expanded legal review to address threats by belligerent nations using economic tools. In 2018, Congress passed the Foreign Investment Risk Review Modernization Act of 2018 (FIRRMA).[253] FIRMMA allows the federal

---

248. *See* Federal Trade Commission Act of 1914, ch. 311, § 5, 38 Stat. 717, 719 (current version at 15 U.S.C. § 45(a)(1) (2018) ("Unfair methods of competition in or affecting commerce, and unfair or deceptive acts or practices in or affecting commerce, are hereby declared unlawful.")).

249. F.T.C. v. Wyndham Worldwide Corp., 799 F.3d 236, 246 (3d Cir. 2015) ("the FTC Act expressly contemplates the possibility that conduct can be unfair before actual injury occurs. 15 U.S.C. § 45(n) ("[An unfair act or practice] causes or is likely to cause substantial injury" (emphasis added)."); *see generally* Luke Herrine, *The Folklore of Unfairness*, 96 N.Y.U.L. Rev. 431, 528 (2021) ("A notion of unfairness devoted to articulating standards of fair dealing rather than figuring out how to ensure that consumers make rational decisions would focus on these structural problems as worthy of redress."); *see also* Matthew W. Sawchak & Kip D. Nelson, *Defining Unfairness in "Unfair Trade Practices"*, 90 N.C. L. Rev. 2033 (2012).

250. *See* Michael Flynn, *"The Lie, the Bigger Lie, and the Biggest Lie"-Unfair and Deceptive Trade Practices of Tripadvisor and Other Online Review Websites*, 36 J.L. & Com. 23, 30 (2017) ("Federal case law defines a deceptive trade practice as an act or practice that has the tendency or capacity to deceive consumers."); *see also* Peter S. Menell, *Regulating "Spyware": The Limitations of State "Laboratories" and the Case for Federal Preemption of State Unfair Competition Laws*, 20 Berkeley Tech. L.J. 1363, 1379 (2005) ("The interactivity of the Internet, in combination with advances in software and database technology, has enabled new forms of advertising that were never before feasible on a wide scale."); *see also* Joel B. Hanson, *Liability for Consumer Information Security Breaches: Deconstructing Ftc Complaints and Settlements*, 4 Shidler J.L. Com. & Tech. 11 (2008).

251. *See* Adams v. Alcolac, Inc., 974 F.3d 540, 543 (5th Cir. 2020) ("Under JASTA, 'liability may be asserted as to any person who aids and abets, by knowingly providing substantial assistance, or who conspires with [a] person who commit[s] . . . an act of international terrorism.'. . . (18 U.S.C. § 2333(d)(2)). 'No action shall be maintained under section 2333,' however, 'for injury or loss by reason of an act of war. § 2336(a).'").

252. *Id.*

253. Foreign Investment Risk Review Modernization Act of 2018, Pub. L. No. 115-232, 132 Stat. 2173 (2018).

government to control foreign interference through investment, making certain transactions subject to review by the Committee on Foreign Investment in the United States (CFIUS).[254] CFIUS has operated since 1975, granting the President authority "to block or suspend proposed or pending foreign "mergers, acquisitions, or takeovers" of U.S. entities, including through joint ventures, that threaten to impair the national security."[255] Under congressional guidance, the U.S. Department of Treasury has promulgated new regulations which expand CFIUS authority. Among the authority granted in 2020 has been authority over transactions involving "TID U.S. businesses," a new acronym for critical technologies, critical infrastructure, and personal data.[256]

The regulations also provide a lengthy definition of sensitive personal data that covers U.S. government and military personnel or contractors, financial data, health care and health status data, geolocation data, biometric and genetic data, stored communications, and more.[257] The scope of the sensitive personal data is sufficiently broad to include essentially all metaverse platform operators as well as most social media services. The only limitation is that the regulation excludes entities that have collected data on one million or fewer individuals, though this limitation will not apply if the entity has the capability to exceed the one-million individual threshold.[258] This same definition could serve to provide a jurisdictional floor. Companies offering these services to fewer than one million users would not be required to submit to the reporting requirements unless other conditions were triggered, such as evidence of actual misuse by terrorist organizations or belligerent nations.

Congress could choose to expand the anti-terrorist duties and disclosure requirements under FIRRMA rather than JASTA. This would send the signal that the goal is sovereign protection rather than tort liability. Another benefit of adding the disclosure regulations to FIRMMA is that it suggests that critical infrastructure and critical TID operations have an affirmative duty to protect their systems from both attacks and systemic abuse. Even if FIRRMA were used as the framework for new disclosure requirements, it would be important that the FTC or another enforcement agency were specifically empowered to assess, and if necessary, bring

254. *See* 31 Fed. Reg. Parts 800 and 801 (Jan. 17, 2020) ("FIRRMA amended and updated section 721 (section 721) of the Defense Production Act of 1950 (DPA), which delineates the authorities and jurisdiction of the Committee on Foreign Investment in the United States (CFIUS or the Committee).").

255. *CFIUS Reform Under FIRRMA*, CONG. RES. SERVICE (Feb. 21, 2020), https://sgp.fas.org/crs/natsec/IF10952.pdf ("CFIUS is an interagency body comprised of nine Cabinet members, two ex officio members, and others as appointed that assists the President in overseeing the national security risks of FDI in the U.S. economy.").

256. See Antonia I. Tzinova, *New CFIUS Regulations Finally Take Effect*, HOLLAND & KNIGHT ALERT (Feb. 13, 2020), https://www.hklaw.com/en/insights/publications/2020/02/new-cfius-regulations-finally-take-effect; *Foreign Investment 2020 (Part 3): CFIUS Spotlight on "TID" U.S. Businesses*, MORRISON & FOERSTER (Oct. 15, 2019), https://www.mofo.com/resources/insights/191015-foreign-investment-2020.html.

257. *See* 31 Fed. Reg. § 800.241.

258. *Id.* at § 800.241(a)(B).

administrative actions against enterprises that were creating opportunities for exploitation by terrorists or belligerent states. These affirmative obligations will be increasingly important if the Web3 ethos of decentralized systems and disaggregated organizations take hold. Decreasing the centralization of economic power is a laudable goal, but platform operations must still meet core security and integrity requirements. Creating affirmative standards on anti-terrorism efforts and mandatory reporting will go a long way to assure that the proper balance is maintained.

VI. CONCLUSION

Since the Nuremberg Trials failed to hold corporate executives responsible for the large-scale sale of poisonous gas to Nazi death camps, international law has expanded its definitions of war crimes and crimes of atrocity. Nonetheless, corporations themselves are not within the jurisdiction of the International Criminal Court. Similarly, domestic criminal laws may not have sufficient reach. Domestic tort law has become the alternative to legal responsibility, but it reaches few of the perpetrators of human rights violations or large-scale war crimes or crimes of atrocity.

As the potential for everyday devices to become agents of government suppression and commercial devices become the tools of autonomous military campaigns, governments, prosecutors, public officials, and consumers must ask questions about which corporate accommodations are acceptable and which should result in culpability for complicity with totalitarian regimes.

The ease with which synthetic media can replace and drown out more carefully sourced legitimate media and AI can be manipulated to promulgate false information as if it originated with thousands of real individuals both require stronger responses by media platforms. In the era of the metaverse, the role of the platform may be harder to identify, and the algorithms or systems by which information is exchanged ever harder for an individual to gauge content accuracy.

The only viable option to stop the most heinous of this systematic misuse is through the creation of new, affirmative duties on the part of large platforms (meaning those with at least one-million customers). The affirmative duty to actively police and remove terrorist content and false content from hostile nations would improve the experience and trustworthiness of the metaverse and media platforms while decreasing the impact of terrorism.

Unless there is an obligation on the part of the companies that make up the communications infrastructure, nothing in JASTA or the ATA will reduce the continuing threat of terrorism. Even this proposal is modest. That is intentional. The immediate goal is to stop the pervasiveness of terrorist materials and reduce the effectiveness of calls for genocide. If these are proven effective, then there might be additional opportunities to expand their scope and further reduce the resources available to terrorist organizations. But this first step is essential to unleash the potential of Web3.

Two decades ago, a student author noted that the new millennium had created an imperative to put aside our hostilities. He wrote, "[a]s the world becomes smaller, easing traditional, seemingly justifiable, hatreds will be one of the most important requirements of the new millennium. Communications media will continue to be an essential factor in bringing this about."[259] Society failed to heed such warnings. It is time to listen once again. Congress cannot legislate love for one's fellow man, but it can do something to slow the expansion of hate. And since it can, it must.

259. Dombeck, *supra* note 195, at 43.