Santa Clara University
School of Law

Santa Clara High Technology Law Journal

Volume 36 | Issue 1 Article 2

2-2-2020

THE IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE IN HARD AND SOFT COUNTERTERRORISM EFFORTS ON SOCIAL MEDIA

Schnader, Jonathan

THE IMPLEMENTATION OF ARTIFICIAL INTELLIGENCE IN HARD AND SOFT COUNTERTERRORISM EFFORTS ON SOCIAL MEDIA

By Jonathan Schnader[1]

The United States government excels at hard measures to counterterrorism, like military operations, non-kinetic information operations, and criminal prosecutions. However, in terms of counterterrorism, the First Amendment obstructs the United States government's efforts to prevent and contain the recruitment efforts of terrorist organizations like ISIS, which have professional-grade social media marketing wings. But, counterterrorism efforts do not need to rest solely within the government. Social media platforms, which can be used in softer counterterrorism approaches, must step up to the plate to combat terrorist recruitment on their platforms. Indeed, some social media platforms have made great strides in deactivating terrorist accounts, but they must continue to evolve to combat and minimize radicalization. This Article proposes that artificial intelligence can aid both the government and the private sector to combat terrorists and terrorist recruitment. Ultimately, the Article proposes that the government use AI to further its hard counterterrorism efforts, and social media platforms should employ AI to augment their soft counterterrorism approaches.

---

[1] Jonathan A. Schnader lives in Washington, D.C. He earned his bachelor's degree in Psychology and Classical Humanities from Miami University of Ohio. He earned his J.D. *cum laude* from Syracuse University College of Law with a Certificate of Advanced Study in National Security and Counterterrorism Law. Following law school, he worked as an Assistant Public Defender in Rochester, NY for five and a half years handling just under four-thousand criminal cases. In addition to being licensed to practice law in New York and Washington D.C., he became a Certified Anti-Money Laundering Specialist. Jonathan recently graduated with distinction from Georgetown University Law Center, where he completed a Master of Laws in National Security Law. His academics and current practice focus on national security dimensions of several areas, including cybersecurity; artificial intelligence; blockchain and cryptocurrency; intelligence and counterintelligence; and social media.

CONTENTS

INTRODUCTION

Over the past ten years, social media platforms like Facebook, YouTube, and Twitter[2] quickly integrated themselves into the daily lives of people across the globe, becoming a central part of everyday life from keeping up with friends, to getting news and current events.[3] The centrality of social media in daily life – specifically for Americans – is apparent through daily use statistics: "Fully 74% of Facebook users say they visit the site daily, with around half (51%) saying they do several times a day."[4] Indeed, the accessibility of social media to everyday, normal people has changed the scope of human interaction, from how people gather information to how people communicate. In terms of American values, the freedom to communicate freely, without government intervention, is a foundational and bedrock principle woven into the fabric of American society by way of the First Amendment to the Constitution:

> A fundamental principle of the First Amendment is that all persons have access to places where they can speak and listen, and then, after reflection, speak and listen once more. . . . While in the past there may have been difficulty in identifying the most important places (in a spatial sense) for the exchange of views, today the answer is clear. It is cyberspace—the 'vast democratic forums of the Internet' in general, and social media in particular."[5]

The guarantees of the First Amendment are so powerful, and the use of social media in society is so ubiquitous, that in *Packingham*, above, the Supreme Court struck down post-conviction prohibitions for convicted sex offenders (perceived by many to be the most contemptible types of criminals) from using social media because to do so would be to:

> bar[] access to what for many are the principal sources for knowing current events, checking ads for employment, speaking and listening in the modern public square, and otherwise exploring the vast realms of human thought and

---

[2] For the purposes of this Article, social media describes traditional social media like Facebook, Twitter, etc., but also includes interactive search engines like Google, and even online retail with interactive qualities, like Amazon.

[3] *See* ELISA SHEARER & KATERINA EVA MATSA, NEWS USE ACROSS SOCIAL MEDIA PLATFORMS 2018 (Pew Research Center, Sept. 10, 2018), https://www.journalism.org/2018/09/10/news-use-across-social-media-platforms-2018/.

[4] AARON SMITH & MONICA ANDERSON, SOCIAL MEDIA USE IN 2018 5 (Pew Research Center, Mar. 1, 2018), https://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.

[5] Packingham v. North Carolina, 137 S.Ct. 1730, 1735 (2017) (quoting Reno v. American Civil Liberties Union, 521 U.S. 844, 868 (1997).

> knowledge. These websites can provide perhaps the most powerful mechanisms available to a private citizen to make his or her voice heard.[6]

But, a side effect of freedom is that sometimes, it is not only used for good, but also for ill. In the case of social media, terrorist networks, specifically the Islamic State in Iraq and Syria ["ISIS"], have turned to social media as their prime recruitment tool: it provides a platform for them to broadcast their message to millions of people; it bestows a degree of anonymity or pseudonymity; and most importantly for their purposes, if they closely manage their speech, it insulates them from investigation or prosecution by their strongest enemy, the American government. "Social networking sites are known for their ability to bring like-minded people together, and terrorist organizations utilize these sites to recruit, fundraise, and spread terrorist propaganda. These websites create a convenient and inexpensive platform for terrorist organizations to expand their global reach, amass support from other like-minded extremists, and capitalize on a larger network of diverse talents and skills."[7]

Platforms, like Facebook, project images of being virtual spaces for pure First Amendment expression. However, some platforms have moved beyond their desultory self-policing plans, into more meaningful paradigms for self-regulation, claiming to have increased their internal methods for parsing violent, extremist, or illegal content. For example, Jack Dorsey, Chief Executive Officer ["CEO"] of Twitter, eloquently stated why violent, malicious, and hateful content on Twitter impairs the free flow of ideas: "Twitter is built and measured by how we help encourage more healthy debate, conversation, and critical thinking. Conversely, abuse, malicious automation, and manipulation detracts from it."[8]

With *billions* of users and *over a billion* hours of content consumed by users each *day*,[9] the ability of a platform to monitor all potentially objectionable material is beyond human comprehension and galactic in scope. Who or *what* is in a position to digest that quantum of information? The answer, unsurprisingly, is artificial intelligence ["AI"]. Similar to how social media changed the global geo-social-

---

[6] *Packingham*, 137 S.Ct. at 1737.

[7] Nicole Phe, *Social Media Terror: Reevaluating Intermediary Liability Under the Communications Decency Act,* 51 SUFFOLK U. L. REV. 99, 100-01 (2018).

[8] *Testimony of Jack Dorsey Chief Executive Officer, Twitter, Inc.*: *Hearing Before the S. Select Comm. on Intelligence*, 116th Cong. 1 (Sept. 5, 2018).

[9] YouTube reports that its over one billion users watch one billion hours of content daily, which equals approximately 114,115 years of content consumed per day. *See e.g.*, *YouTube for Press*, YOUTUBE, https://www.youtube.com/yt/about/press/ (last visited Aug. 17, 2019).

political topography, so too will AI contribute to and/or cause a rapid shift in how individuals, companies, and governments synthesize and understand information and data, in local, regional, national, and global contexts. The current presidential administration, acknowledging how pivotal AI is, issued an executive order declaring that the government will go "full steam ahead" to encourage and foster AI development both in government and in the private sector.[10]

For purposes of this discussion, the definition of "AI" must be established to keep a uniform vocabulary throughout this Article. Therefore, "AI," for the purposes of our discussion, means "systems that can emulate, augment, or compete with the performance of intelligent humans in well-defined tasks."[11] This broad definition includes all types of AI systems, from simpler algorithms (narrow and/or "dumb" AI systems) to Artificial General Intelligence, defined as "'strong' [AI] with the full range of cognitive capacities typically possessed by humans, including self-awareness."[12]

How can AI systems help the U.S. counterterrorism effort? Generally, scholarship since September 11, 2001 focuses on a spectrum of counterterrorism strategies that range from "hard" strategies like covert action, targeted killings, war, and prosecution, to "soft" counterterrorism strategies focusing on radicalization prophylaxis, deterrence, and rehabilitation. This analysis seeks to elucidate the benefits of AI applied to this spectrum of counterterrorism strategies.

Road Map

First, this analysis will discuss counterterrorism terminology, strategy, and scholarship. This analysis will then summarize various AI fundamentals. Following the description of AI systems, this analysis will survey methods for employing AI in strategic counterterrorism efforts from the perspective of the U.S. government. After discussing the use of AI by the government, it will analyze the use of AI systems by the private sector, particularly social media platforms in combatting terrorism. Finally, this Article will conclude with some general guidance and proposals. In sum, AI is best suited for the U.S.

---

[10] *See* Exec. Order No. 13859, 3 C.F.R. § 3967 (Feb. 11, 2019), https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.

[11] Shannon Vallor & George A. Bekey, *Artificial Intelligence and the Ethics of Self-Learning Robots*, *in* ROBOT ETHICS 2.0: FROM AUTONOMOUS CARS TO ARTIFICIAL INTELLIGENCE 339 (Patrick Lin, Keith Abney, & Ryan Jenkins eds., 2017). This book contains some of the best and most accessible descriptions, summaries, and explanations of AI systems, and I have cited to their definitions in several other AI related articles.

[12] *Id.* at 339-340.

government for use in "hard" approaches to counterterrorism, and excellent for use in "soft" approaches to counterterrorism employed by social media platforms.

IV. CONCEPTUAL GROUNDINGS IN COUNTERTERRORISM

The United States underwent an extreme transition in national security strategy after the September 11th attacks, mobilizing for war and shifting to an unrelenting hunt for terrorists. After what most consider to be a successful campaign against Al Qaeda, a new terrorist organization emerged on the global stage: ISIS. ISIS proclaimed that it established a new world order and Islamic caliphate in the Middle East, seizing territory in warn-torn Syria and Iraq, establishing a fledgling government and recruiting people to its cause on social media. Its deft use of social media as a recruitment tool resulted in thousands of people from all over the globe traveling to fight on behalf of ISIS, as well as radicalizing and committing acts of violence in their home countries in the name of ISIS.[13]

The counterterrorism strategy spectrum ranges from "hard" approaches which "are more militaristic in approach–involving targeted assassinations or even warfare," to "soft" counterterrorism "programs [which] seek to undo the radicalization process by engineering the individual's return to moderate society, usually by providing them with [a] stable support network, probing their original reasons for radicalizing, and divorcing them from their extreme beliefs and social contacts."[14] The scholarship evaluating potentially effective counterterrorism strategies seems to endorse a holistic approach: a State should employ hard measures – investigation, prosecution, and incarceration – likewise the U.S. government should use soft measures to deter and prevent radicalization before it starts, and reverse the radicalization process as it progresses.

This section will contextualize terrorist recruitment on social media in general. This analysis will then move to discuss counterterrorism measures, briefly mentioning "hard" approaches, but ultimately focusing on "soft" measures like counter-radicalization. As discussion later in the analysis will demonstrate, both "hard" and "soft"

---

[13] Importantly, the threat of "domestic terrorists," deserves scholarship and analysis. This analysis focuses mostly on terrorist recruitment by ISIS because vast scholarship and data exist about ISIS's methods. Many of the approaches and challenges discussed in this analysis overlap with domestic terrorist threats.

[14] ELLIE B. HEARNE & NUR LAIQ, A NEW APPROACH? DERADICALIZATION PROGRAMS AND COUNTERTERRORISM 3 (International Peace Institute, June 2010), https://www.ipinst.org/wp-content/uploads/publications/a_new_approach_epub.pdf [hereinafter IPI Report].

approaches to counterterrorism stand to benefit from AI systems tailored to their objectives.

A. Terrorist Recruitment

Recently, former Federal Bureau of Investigation (FBI) Director, James Comey, highlighted the burgeoning surge of terrorist use of social media to recruit individuals for their cause: "During my time as director we had an explosion of the use of social media to proselytize by terrorist organizations, especially the so-called Islamic State, to recruit and to direct terrorist activity through the internet. That was a big change."[15] As the FBI Director at the time, James Comey bore responsibility for adapting Federal law enforcement and counterintelligence resources to combat the emerging threat of social media in terms of terrorist recruitment:

> The technology platforms are where we live, today. It's where, depressingly, increasingly our social lives – not mine – but normal people's social lives are. That, in a cool way and in a disturbing way, it's made everybody a next-door neighbor to everybody else. And, so there is a lot great about that, but there is a lot of danger in that because it brings crazy people and dangerous people close to each other, and allows them to use those platforms to recruit, to share information and plans, to direct, to control.[16]

Social media serves as a worldwide "soapbox" and recruitment platform: "social networking allows terrorists to reach out to their target audiences and virtually 'knock on their doors'–in contrast to older models of websites in which terrorists had to wait for visitors to come to them."[17]

More specifically, ISIS:

> . . . has employed social media to gain the attention of mass-media and strategic audiences, inflate and control its messaging in support of its narrative in order to recruit and radicalize followers, deter their opponents and to raise funds. . . . At the same time, they manage to construct their 'self-expression' in a way supportive of their narrative, while displaying an understanding of how to disrupt an opponent's narrative and online activities by exploiting their messaging

---

[15] *Bonus Edition: James Comey at Verify 2019*, LAWFARE PODCAST (Apr. 11, 2019), https://www.lawfareblog.com/lawfare-podcast-bonus-edition-james-comey-verify-2019.

[16] *Id.*

[17] Gabriel Weimann, *New Terrorism and New Media*, WILSON CENTER COMMONS LAB RESEARCH SERIES VOL. 2, 2014 at 3, http://www.wilsoncenter.org/publication/new-terrorism-and-new-media.

> in order to position themselves and their 'brand' amongst other jihadist factions in the Middle East.[18]

Recent estimates suggest that ISIS recruited over 16,000 to join their cause through the use of social media.[19] Moreover, the ISIS threat of violence does not consist of only people who physically join ISIS in the Middle East, but also "homegrown violent extremists," who are radicalized through social media to commit acts of violence in their home countries.[20] Some officials believe that the threat of violence from "homegrown violent extremists" is the "most likely and immediate threat" to the United States.[21] In his 2018 Global Threat Assessment, Director of National Intelligence, Dan Coats, specifically pointed to ISIS – as well as homegrown violent extremists – as "continuing terrorist threats to US interests and partners worldwide."[22] In the most recent Global Threat Assessment, Coats noted the despite major territorial losses suffered by ISIS, if counterterrorism pressure exerted by the United States and its partners were to wane, ISIS could rebuild "key capabilities," including "external operations" and "media production."[23]

Terrorist organizations, often using Facebook or Twitter, seek out "disenfranchised or disaffected people by tweeting, retweeting, and using popular hashtags . . . relating to divisive current events"; "create [an] online micro-community around the targeted recruit"; and "encourage the recruit to isolate himself from moderating influences."[24] As an individual becomes more radicalized, the recruiters shift their communications to private social media platforms

---

[18] Thomas Elkjer Nissen, #TheWeaponizationOfSocialMedia 58-59 (Royal Danish Defence College, 2015) (citing Mark Borkowski, *Isis and the Propaganda War: How the social-savvy Extremist Are Dominating the Headlines*, THE DRUM (June 25, 2014), https://www.thedrum.com/opinion/2014/06/25/isis-and-propaganda-war-how-social-savvy-extremists-are-dominating-headlines.

[19] Susan Klein & Crystal Flinn, *Social Media Compliance Programs and the War Against Terrorism*, 8 HARV. NAT'L SEC. J. 53, 65 (2017).

[20] *Id.*

[21] *Id.* (citing *Current Terrorist Threat to the United States: Hearing Before the S. Select Comm. on Intelligence*, 114th Cong. (2015) (statement of Nicholas J. Rasmussen, Director, Nat'l Counterterrorism Center)) (internal quotations omitted).

[22] Dan Coats, *Statement for the Record: Worldwide Threat Assessment of the US Intelligence Community*, Director of Nat'l Intelligence, S. Select Comm. on Intelligence (Feb. 13, 2018) at 9.

[23] Dan Coats, *Statement for the Record: Worldwide Threat Assessment of the US Intelligence Community*, Director of Nat'l Intelligence, S. Select Comm. on Intelligence (Jan. 29, 2019) at 11.

[24] Klein & Flinn, *supra* note 19, at 66 (quoting J.M. Berger, *Tailored Online Interventions: The Islamic State's Recruitment Strategy*, COMBATING TERRORISM CTR. SENTINEL, Oct. 2015 at 21), https://ctc.usma.edu/tailored-online-interventions-the-islamic-states-recruitment-strategy/ (internal quotations omitted).

with encryption like WhatsApp, Kik, Telegram, etc., and this transition from the public to private space is known as "going dark."[25]

ISIS' uniquely insidious, "comprehensively designed and carefully branded operation"[26] that uses professional level marketing tactics, differs starkly from Al Qaeda's notoriously low-budget, grainy VHS recordings of clerics speaking into a camera. Indeed, "[ISIS] is as much a media conglomerate as a fighting force."[27] A 2015 think tank reported that ISIS releases "38 new items per day–20-minute videos, full-length documentaries, photo essays, audio clips, and pamphlets, in languages ranging from Russian to Bengali."[28] ISIS evidently embraces a strategy of mass-appeal, using numerous narratives (utopia, mercy, brutality, war, belonging, *inter alia*) to further its goals:

> [ISIS's] supporters are presented with a comprehensive idea of what life in its 'caliphate' is like. Brutality plays an important role in this image, but is by no means the key to [ISIS's] appeal, as is regularly argued. Mercy and belonging, both narratives that featured heavily . . . are no longer as prominent as they had been in the past, something that is reflective of a shift in priorities for the group. Far more conspicuous are its attempts to reinforce the victimhood narrative by playing upon the 'War on Islam' . . . [ISIS's] propagandists expend a great amount of effort portraying it as militarily dynamic and ever-expanding. After all, the perception of momentum is central to its 'winner's messaging.' The most prominent narrative, by far, is that of utopia. At once the most appealing promise of the group . . . [ISIS's] media strategists ensure that utopia is sold as a comprehensive project, where the economy flourishes, 'Islam' is implemented, wildlife thrives, rule of law prevails and the government governs.[29]

In sum, the daily ISIS propaganda blast coupled with recruiters on social media function to radicalize numerous individuals.

---

[25] Klein & Flinn, *supra* note 19, at 67.

[26] CHARLIE WINTER, DOCUMENTING THE VIRTUAL CALIPHATE 39 (Quilliam Foundation, Oct. 2015) http://www.quilliaminternational.com/wp-content/uploads/2015/10/FINAL-documenting-the-virtual-caliphate.pdf [hereinafter Quillam Report].

[27] Brendan I. Koerner, *Why ISIS is Winning the Social Media War*, WIRED (Apr. 2016), https://www.wired.com/2016/03/isis-winning-social-media-war-heres-beat/.

[28] *Id.* (citing the Quilliam Report, *supra* note 26, at 25).

[29] Quilliam Report, *supra* note 26, at 38-39.

B. Hard Approaches to Counterterrorism

The term "hard" refers to an approach to counterterrorism based on "coercion (such as military force and economic sanctions) . . . ."[30] As discussed above, "[c]ontributing significantly to [terrorist attacks connected with or inspired by ISIS] is the risk that individuals will become radicalized by viewing extremist material on the Internet. While this threat was growing many years before [ISIS], [ISIS's] significant online presence makes this a particular challenge in the current security environment."[31] Hard approaches – like military intervention, covert action, economic sanctions, and criminal arrests and prosecutions – play an important part to combatting counterterrorism. Primarily, the responsibility to implement hard counterterrorism measures lies with the government. The nexus between hard counterterrorism efforts and social media exist in two major buckets: (1) using social media as a source of intelligence for military or covert operations abroad and (2) criminalizing behavior on social media aimed at radicalization or recruitment.

The U.S. military uses "kinetic" and "non-kinetic" approaches to bolster its hard approaches to counterterrorism: "kinetic" refers to "[a] proactive and aggressive approach, kinetic action targets enemy combatants and their supporters to neutralize, capture or eliminate them."[32] Considered the most aggressive approach, it involves "the removal of central nodes or brokers, or they involve the breaking of key ties or links among individuals, groups, or organizations."[33] In other words, by analyzing various terrorist groups or networks, military officials can choose the most integral connections and disrupt them through person-targeting or group-targeting.[34] On the other hand, "non-kinetic" measures are milieu approaches in nature. "Non-kinetic" measures involve "a more subtle and patient application of power by seeking to undermine terror networks 'more through cooperation and collaboration with partners than through unilateral American action, more with the diplomatic and economic tools of national power than

---

[30] Keiran Hardy, *Hard and Soft Power Approaches to Countering Online Extremism*, GRIFFITH CRIMINOLOGY INST., Jan. 2017, at 2, https://www.researchgate.net/publication/326288572_Hard_and_soft_power_approaches_to_countering_online_extremism.

[31] *Id.* at 1.

[32] Nancy Roberts & Sean F. Everton, *Strategies for Combating Dark Networks*, 12 J. OF SOC. STRUCTURE 1, 4 (Jan. 5, 2011), https://www.cmu.edu/joss/content/articles/volume12/RobertsEverton.pdf.

[33] *Id.*

[34] *See id.*

with the military, stressing inspiration rather than prescription.'"[35] Non-kinetic strategies include institution building (civil assistance, and humanitarian aid); psychological operations ("dissemination of information for the purpose of influencing the emotions, perceptions, attitudes, objective reasoning, and ultimately the behavior of foreign nationals"); and information operations ("electronic warfare and computer network operations").[36] Military approaches (both kinetic and non-kinetic) cost the U.S. government and taxpayers significant amounts of money,[37] present a high risk to U.S. servicepersons, and implicate international legal principles[38] and foreign policy.

Domestically, hard counterterrorism measures arise in terms of criminal prosecutions. The U.S. menu of terrorist related crimes,[39] coupled with other law enforcement tools, allow prosecutors to bring charges and try cases against terrorists in U.S. courts. Courts may impose lengthy sentences on those terrorists upon conviction. When it comes to criminalizing terrorist recruitment online, however, criminalization of radical speech or dissemination of distasteful terrorist aligned materials abuts or overlaps with cherished First Amendment rights (discussed *infra*), making the U.S. government involvement in censorship, regulation of social media, or arrests for terrorist-aligned speech worrisome.[40]

C. Soft Approaches to Counterterrorism

Before further discussing the nuances of counterterrorism strategies on the "softer" end of the spectrum, it would be useful to

---

[35] *Id.* at 5.

[36] *Id.* at 6.

[37] *See* NETA C. CRAWFORD, UNITED STATES BUDGETARY COSTS OF THE POST- 9/11 WARS THROUGH FY2019: $5.9 TRILLION SPENT AND OBLIGATED (Brown U. Watson Inst. Of Int'l and Pub. Aff., Nov. 14, 2018), https://watson.brown.edu/costsofwar/files/cow/imce/papers/2018/Crawford_Costs%20of%20War%20Estimates%20Through%20FY2019.pdf.

[38] For a comprehensive discussion of the international legal landscape and the so-called "War on Terror," *see e.g.*, HELEN DUFFY, THE 'WAR ON TERROR' AND THE FRAMEWORK OF INTERNATIONAL LAW (Cambridge U. Press, 2d ed. 2015).

[39] *See e.g.*, 18 U.S.C. §2331 *et seq*. Notably, the U.S. code lacks a domestic terrorism criminal statute. For a discussion about a potential domestic terrorism statute, *see* Mary B. McCord, *It's Time for Congress to Make Domestic Terrorism a Federal Crime*, LAWFARE (Dec. 5, 2018), https://www.lawfareblog.com/its-time-congress-make-domestic-terrorism-federal-crime.

[40] Hardy, *supra* note 30, at 4-8 (analyzing the United Kingdom's efforts to criminalize terrorist behavior facilitated by the Internet and social media, proscribing the act of downloading and/or printing extremist material, as well as posting statements in online fora encouraging or supporting terrorism. Clearly, such proscriptions would run afoul of First Amendment protections of free speech and would thus be unacceptable in United States domestic law contexts). *See e.g.*, U.S. CONST. amend. I ("Congress shall make no law . . . abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble . . .").

define terms. "Anti-radicalization" refers to measures used to prevent radicalization of individuals before it starts; "counter-radicalization" refers to stopping the radicalization process once it has begun with a specified person; and "deradicalization" refers to mitigating or "un" radicalizing a person who has completed the radicalization process.[41] This analysis will use the term "counter-radicalization" to also refer to the general effort of combating radicalization as a policy consideration.

In a counterterrorism context, "soft" refers to the "attractive power of a nation's culture and values"[42] in curbing terrorist behavior. These soft approaches include prison deradicalization programs; aftercare programs that monitor and remind individuals about the importance of distancing him/herself from radicalization triggers; community involvement in anti- and counter-radicalization, as well as post-institutionalization deradicalization rehabilitation programs; financial incentives; *inter alia*.[43] Saudi Arabia began its holistic counterterrorism program in 2005, and in addition to its strong criminal laws proscribing extremism and terrorism, it employs manifold soft measures that focus substantially on counter-radicalization and deradicalization.[44] The Saudi initiative does indeed fit the definition of "deradicalization": it begins "with the suspected terrorist's arrest," after which the "individual is immediately engaged in dialogue."[45] Participants, should they decide to partake in the "'religious, psychological, and cultural'" program,

> . . . are allowed to attend social events, including family gatherings and sports (often games among [participants], police, and program tutors to foster trust), but are engaged all the while in a program consolidating the 'correct notions and concepts' of Islam. The program also encourages participants to marry (with financial support), and to pursue further education. In returning the [participants] to jobs they held prior to radicalizing, the program seeks to ground repentant extremists in a stable environment.[46]

---

[41] *See* Lindsey Clutterbuck, *Deradicalization Programs and Counterterrorism: A Perspective on the Challenges and Benefits*, MIDDLE EAST INST., (June 10, 2015), https://www.mei.edu/publications/deradicalization-programs-and-counterterrorism-perspective-challenges-and-benefits. Notably, Dr. Clutterbuck's definition of deradicalization specifically excludes persons who are completely radicalized but do not act on their beliefs. Consequently, Dr. Clutterbuck's definition of "deradicalization" only applies those persons caught by the criminal justice system.

[42] Hardy, *supra* note 30, at 2 (internal quotations omitted).

[43] *See generally*, IPI Report, *supra* note 14.

[44] *See id.* at 7.

[45] *Id.*

[46] *Id.*

The program participants may receive healthcare and financial assistance, not only for themselves but for their families, because "family loyalty seemed to be stronger than any loyalty to the state, and so focusing on families seemed to generate longer-lasting results and fewer regressions to patterns of violence."[47]

The Saudi program's most unique feature, *Sakina*, focuses resources combating radicalization in online fora: "[a] carefully appointed group of intellectuals visit websites where radicals congregate online and they challenge extreme interpretations of Islam. They also carry out youth dialogues over the Internet, mirroring the use of the Internet by violent extremists to recruit prospective terrorists."[48] As noted above, a *Sakina*-type anti/counter/de-radicalization initiative would encounter substantial legal friction in the United States. One reason for this friction being that it requires endorsement by the government of a "correct" view of Islam, breaching mandates that the government not interfere in matters of religion under the Establishment Clause.[49]

The increase of terrorist presence on social media requires a response. The question, however, is whether hard or soft measures will be best suited to the task, considering the particular legal challenges of regulating social media in the United States. As will be discussed below, the advantages of AI systems in processing voluminous amounts of data should be channeled into countering terrorism by using social media. Before discussing where AI best fits into the U.S. counterterrorism strategy, it is essential to describe *how* AI works in order to understand its strengths and weaknesses.

V. DISCUSSION OF AI FUNDAMENTALS

A. Defining Artificial Intelligence

At least some AI experts consider AI to be systems "that contain[] machine learning [] and deep learning []," and "a combination of reinforcement learning [] and deep learning . . . ."[50] Therefore, understanding the challenges facing AI in combatting terrorist

---

[47] *Id.* at 8.

[48] *Id.* at 8-9.

[49] "Congress shall make no law respecting an establishment of religion, *or prohibiting the free exercise thereof*," U.S. CONST. amend. I (emphasis added).

[50] Oludare Isaac Abiodun et al., *State-of-the-art in artificial neural network applications: A survey*, HELIYON, Nov. 2018, at 8, 10, https://reader.elsevier.com/reader/sd/pii/S2405844018332067?token=BC6FC8B560036BB840A960796CEF64F4711376FB7E28B5A2D085BC559B646EBA33DCFAEF3EF636AB8ACE5EC250891ED0 [hereinafter ANN Survey Article].

recruitment requires a cogent but simplified discussion of *how* AI systems interpret data, synthesize it, learn from it, and then act autonomously.

What makes AI systems useful to humankind? "In some cases, their value may come from being cheaper, faster, or easier to deploy at scale relative to human expertise."[51] Some areas in which AI systems excel, as noted by the Center for a New American Security, include: data classification; anomaly detection; prediction; and optimization (like improving energy efficiency), at least in part due to skills like faster-than-human reaction times; superhuman precision and reliability; superhuman patience and vigilance; and operations without connections to humans.[52] Thus, AI systems can parse millions of pieces of data in extremely short amounts of time and make judgments or decisions classifying or categorizing that data. How AI systems actually achieve such remarkable results depends on the type of AI system utilized.[53]

B. Machine Learning

A popular type of algorithm that guides AI development is known as a "machine learning" algorithm, and "[t]hese kinds of programs have been around long enough to run unremarkably in the background of everyday US life."[54] "Given a goal, learning machines adjust their behavior to optimize their performance to achieve that goal."[55] Machine learning algorithms are:

> processes capable of learning from data to make ever-more accurate decisions and predictions. Given a set of salient features to look for (for example, what distinguishes a cat image from a dog image) and a mass of data on which to train (a bunch of labeled cat and dog images), a machine-learning algorithm can come to recognize those features in new data (like an unlabeled picture of a domestic quadruped) and draw relevant conclusions (*cat* or *dog*).[56]

---

[51] PAUL SCHARRE ET AL., ARTIFICIAL INTELLIGENCE, WHAT EVERY POLICYMAKER NEEDS TO KNOW 9 (Center for a New American Security, June 19, 2018), https://www.cnas.org/publications/reports/artificial-intelligence-what-every-policymaker-needs-to-know.

[52] *Id.* at 9-10.

[53] *See id.* at 3.

[54] John Fletcher, *Deepfakes, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance*, 70 Theatre J. 455, 458 (Dec. 2018), https://muse-jhu-edu.proxygt-law.wrlc.org/article/715916/pdf.

[55] SCHARRE ET AL., *supra* note 51, at 5. CNAS produces incredible and important policy guidance for cutting-edge issues in national security, particularly the realm of AI.

[56] Fletcher, *supra* note 54, at 458 (emphasis in original).

Some social media platforms already use machine learning to connect users with "relevant" content:

> The algorithms that suggest the next YouTube videos to watch or that flag spam in your email inbox, for example, exemplify machine learning. The more input you give those systems (liking a video, marking a message as spam), the more accurately they 'learn' and successfully predict your preferences.[57]

There is also "reinforcement learning," which "enables a system or an agent [to] learn from the previous experiences gain[ed] in the environment through interaction and observing the results of these interaction . . . [which] helps to mimic or imitate[] the basic pattern in which humans and animals learn."[58]

C. Artificial Neural Networks & Deep Learning

Among the most advanced methods for AI development is the "artificial neural network" ("ANN"), which "can be comparable [to a] machine produced to function the same way the human brain performs a given task of interest."[59] The demonstrable increase in use of ANNs is a testament to its diverse applicability:

> ANNs [have] seen massive use in specific domains, such as[:] diagnosis of hepatitis; speech recognition; recovery of data in telecommunications from faulty software; interpretation of multi-language messages; three-dimensional object recognition; texture analysis; facial recognition; undersea mine detection; and hand-written word recognition. Thus, ANNs can learn by example like people. In some cases, ANNs can be designed for a specific application like data classification or pattern recognition through the learning process.[60]

People often refer to neural networks in conjunction with another buzzword, "deep learning," an advanced type of ANN.[61] "Deep learning" refers to ANNs with complex, hidden layers, meaning the processes that enable the algorithm to make a conclusion based on an input are hidden from its human overseers.[62] To contrast with simple

---

[57] *Id.*

[58] ANN Survey Article, *supra* note 50, at 10.

[59] *Id.* at 4.

[60] *Id.* at 5.

[61] *See e.g.*, Bernard Marr, *Deep Learning Vs. Neural Networks – What's the Difference?,* BERNARD MARR & CO., https://bernardmarr.com/default.asp?contentID=1789 (last visited Aug. 24, 2019).

[62] Fletcher, *supra* note 54, at 459.

machine learning techniques, "[w]hereas machine-learning algorithms require the features they look for in data to be pre-set, deep-learning neural net[works] can determine and detect salient features on their own."[63] Deep learning processes represent a leap forward in AI system autonomy and development.

In general, AI system designs, at this moment in technological development, perform well by analyzing vast sets of data and identifying patterns at light speed after some degree of training. Therefore, the question arises, how do we best apply AI systems to achieve U.S. counterterrorism objectives? The next section will evaluate the best possible ways for the U.S. government and private sector, respectively, to actualize counterterrorism goals with AI systems.

III. GOVERNMENT-BACKED ARTIFICIAL INTELLIGENCE COUNTERTERRORISM STRATEGIES

The U.S. government excels at achieving military dominance on the world stage. While the costs of the "War on Terror"[64] reaches exorbitant levels, the U.S. military campaigns abroad, in many ways, successfully dismantle terrorist organizations. While the government's hard measures accomplish arguably high levels of success, soft approaches to counterterrorism prove to be difficult to implement. After a general discussion of legal hurdles to U.S. holistic counterterrorism strategies, this analysis opines that the government should withdraw from soft counterterrorism measures and focus on the approaches on the hard side of the continuum. It is in that context – military and intelligence – that the government should wield AI.

A. Generally

After the events on September 11th, the U.S. government national security interest in combating terrorism raced to the forefront of U.S. policy. It maintained that primacy for years, culminating in the killing of Osama Bin Laden. However, notwithstanding the evident victory over Al Qaeda, other terrorism-related threats emerged, including ISIS. Indeed, the "hard" counterterrorism methods employed by the U.S. government achieved real results in the "War on Terror," but they certainly did not suffice to prevent the spread of radicalization on social media. As ISIS recruitment proliferated, the U.S. government found

---

[63] *Id.*

[64] The War on Terror campaign was launched by President George W. Bush in response to the September 11th attacks in 2001. The purpose of the War on Terror was to eliminate international extremist and terrorist threats to U.S. interests and the interests of its partners.

itself, constrained by the First Amendment, standing by and brainstorming ways to act, while ISIS recruited thousands of people. Since the threat of radicalization infiltrated the Internet, the United States has grappled with the best method for bringing holistic counterterrorism measures to the social media space.

B. Challenges

The U.S. government faces two major challenges in creating a holistic counterterrorism strategy. First, criminalizing speech or expression on social media treads on First Amendment free speech protections. Second, effective soft measures, like counter-radicalization, will often run afoul of the Establishment Clause under the First Amendment.

1. First Amendment Concerns

Social media platforms must deal with the antagonism between regulation of the social media space and the protections of the First Amendment, which guards against, in relevant part, the abrogation of speech. The effects of the First Amendment protections on social media manifest through individuals using social media as a virtual megaphone for their opinions, thoughts, beliefs, observations, and ideas. The broadcast of all this information is not limited to U.S. persons; indeed, literally billions of people around the globe post content on social media platforms. The sheer volume of such posts collectively by users, the speed at which the content is released and/or consumed, and the visibility of the information to all other users on the platforms, challenges how social media information fits into historical notions of traditional media. Moreover, the public nature of the social media platforms allows access to average people and malign actors alike, including hackers, terrorists, foreign intelligence services, etc.

Although extremist viewpoints tend to fall under the purview of the First Amendment, the protections are limited. Once speech crosses into true threats, fraud, child pornography, libel, incitement, defamation, or imminent threats, the government has the power to prevent and/or criminalize it.[65] So, in considering speech advocating for violence and terrorism, the "clear and present" danger test[66] applies:

---

[65] *See* United States v. Alvarez, 567 U.S. 709, 717 (2012) (striking down the Stolen Valor Act, as a content-based restriction, violated the First Amendment, where defendant pleaded guilty to falsely claiming that he had received the Medal of Honor, reasoning that more than just a lie is required to proscribe untrue speech, because "[t]he remedy for speech that is false is speech that is true"). *Id.* at 727.

[66] *See* Schenk v. United States, 249 U.S. 47, 52 (1919).

> Speech advocating the use of force or crime can only be proscribed where (1) the speech is 'directed to inciting or producing imminent lawless action' – a requirement of intent; and (2) the advocacy is also 'likely to produce such action.' Importantly, when the Court examines the strength of the government interest proffered today, it 'unmistakably insists that any limit on speech be grounded in realistic, factual assessment of harm.'[67]

So long as their public posts refrain from imminent threats, ISIS sympathizers musing about the downfall of the United States may abstractly hide behind the aegis of the First Amendment.[68]

Notably, it is unlikely that the umbra of protection afforded by the First Amendment extends to non-U.S. persons living abroad: interpreting what "the people of the United States" means for several rights in the Constitution, the Supreme Court stated in dicta that "[w]hile this textual exegesis is by no means conclusive, it suggests that 'the people' protected by the Fourth Amendment, and by the First and Second Amendments . . . refers to a class of persons who are part of a national community or who have otherwise developed sufficient connection with this country to be considered part of that community."[69] The lack of extraterritorial attachment of First Amendment rights may give flexibility to the government when operating abroad, especially if it targets non-U.S. persons' use of social media accounts abroad.

Should the government attempt to criminalize abstract, non-imminent violent or extreme speech as a hard counterterrorism measure? The courts would likely strike such legislation down. In that sense, the government faces a challenge virtually unique to the United

---

[67] Tompros et al., *The Constitutionality of Criminalizing False Speech Made on Social Networking Sites in a Post-Alvarez, Social Media-Obsessed World*, 31 Harv. J.L. & Tech. 64, 93 (2017) (citing Brandenburg v. Ohio, 395 U.S. 444, 447 (1969); United States v. Williams, 553 U.S. 285, 321-22 (2008) (Souter, J., dissenting)).

[68] *See* Elonis v. United States, 135 S.Ct. 2001 (2015) (overturning defendant's conviction for making threats over interstate commerce, holding that, where defendant made numerous violent posts about his ex-wife on Facebook, the jury instruction given by the Court, using the reasonable person standard, was "inconsistent with the conventional requirement for criminal conduct – awareness of some wrongdoing," but not considering defendant's first amendment claims (internal quotations omitted)). *Id.* at 2003.

[69] United States v. Verdugo-Urquidez, 494 U.S. 259, 265 (1990) (rejecting defendant's argument that Fourth Amendment protections applied to him, a foreign national, holding "the Fourth Amendment has no application," reasoning that "[a]t the time of the search, he was a citizen and resident of Mexico with no voluntary attachment to the United States, and the place searched was located in Mexico"). *Id.*at 274-75. *See also* United States ex rel. Turner v. Williams, 194 U.S. 279, 292 (1904) (finding that an excludable alien does not enjoy First Amendment rights, reasoning that "[h]e does not become one of the people to whom these things are secured by our Constitution by an attempt to enter forbidden by law.").

States. Other nations that are aligned philosophically in many ways with the United States (like the United Kingdom) have no limitations proscribing extremist speech as a part of their counterterrorism approach, and those prohibitions do not run afoul of their keystone legal principles.[70]

2. Establishment Clause Concerns

The Supreme Court "has been inconsistent about the tests used to analyze the Establishment Clause," and therefore the jurisprudence is not simple.[71] The *Lemon* test uses three prongs to evaluate whether government activity violates the Establishment clause, failure to satisfy any one of the prongs results in a violation[72]:

> the original test included (1) whether the government policy has a legitimate secular purpose; (2) whether the policy's primary effect is one of advancing or inhibiting religion; and (3) whether the policy creates excessive government entanglement with religion.[73]

Crucially, most questions arising under the Establishment Clause deal with endorsement of a particular religion rather than express disapproval of religious beliefs.[74] Most obviously, any attempt by the government to proscribe specific forms or sects of Islam, for example, would violate the Establishment Clause.[75]

Likewise, when it comes to soft approaches like counter-radicalization, the government likewise confronts difficulties. "[C]ounter-radicalization puts the government in the position, vis-à-vis Islam, of serving as a kind of official theologian, taking positions on the meaning of inevitably contested religious concepts and weighing in on one side of debates that rage within a particular faith tradition."[76] In other words, the Establishment Clause prohibits the government from endorsing one religious perspective. So, if the government extols a "mainstream" version of Islam or works to persuade a person online

---

[70] *See* Hardy, *supra* note 30, at 4-8.

[71] Allison Hugi, *A Borderline Case: The Establishment Clause Implications of Religious Questioning by Government Officials*, 85 U. CHI. L. REV. 193, 197 (Jan. 2018).

[72] Edwards v. Aguillard, 482 U.S. 578, 583 (1987).

[73] Hugi, *supra* note 71, at 198-99 (citing Lemon v. Kurtzman, 403 U.S. 602, 612-13 (1971)).

[74] *See* Everson v. Board of Education of Ewing TP et al., 330 U.S. 1 (1947).

[75] *See e.g.*, Awad v. Ziriax, 670 F.3d 1111, 1131 (10th Cir. 2012) (striking down a proposed Oklahoma state constitutional amendment banning Sharia law in Court, applying the different *Larson* strict scrutiny test, reasoning that "[e]ven if the state could identify and support a reason to single out and restrict Sharia law in its courts, the amendment's complete ban of Sharia law is hardly an exercise of narrow tailoring.").

[76] Samuel J. Rascoff, *Establishing Official Islam? The Law and Strategy of Counter-Radicalization*, 64 STAN. L. REV. 125, 162 (2012).

that radical Islamic viewpoints fail to really understand the Koran, the government official crosses the boundary into constitutionally restricted government endorsements of religion.[77]

From a policy perspective as opposed to a legal one, government led counter-radicalization measures, specifically those aimed at Muslims, may isolate Muslims in those communities by treating them differently than other Muslims, and differently than the non-Muslim population. The Saudi *Sakina* program, discussed above, did not struggle in the same fashion for several reasons. One reason being that Saudi Arabia is a Muslim nation where religion, law, and government exist inextricably. In addition to these legal hurdles, from a more practical perspective, people do not consider the U.S. government to be a credible messenger in general.[78]

Considering the legal challenges, if the U.S. government programmed an AI system to, for example, search social media platforms for extremist groups or discussions and then generate automated responses arguing a moderate interpretation of Islam, then this would likely be unconstitutional and would also be a misplacement of resources. Therefore, the government should use AI in a space in which it dominates and excels.

C. Current Approaches and Strategies

AI systems, deployed in a thoughtful way for the government counterterrorism strategy, have the potential to augment the national security effort, particularly abroad in support of "hard" measures. The U.S. military already uses automated weapon technology to bolster its capabilities, but automated weapons are outside this discussion.[79] The U.S. government should (if it does not already) deploy AI systems to provide support to military kinetic and non-kinetic objectives including, but not limited to, locating terrorists, strategizing and prioritizing target individuals in a network, and studying and predicting terrorist movements.

As discussed above, for AI to reach its peak efficacy, it requires vast quantities of data. If a machine learning or deep neural network AI system could access data on terrorist behavior to sift through data and develop patterns, locations, phone numbers, etc., the results could be

---

[77] For deep analyses of Establishment Clause doctrine and counter-radicalization, *see generally* Rascoff, *supra* note 76.

[78] *See e.g.*, BIPARTISAN POLICY CENTER, DIGITAL COUNTERTERRORISM, FIGHTING JIHADISTS ONLINE (Mar. 2018). The U.S. government's counter-messaging program actually caused the attitudes regarding the U.S. government of foreign Arab students to degrade.

[79] For a full discussion of automated weapons systems, *see* PAUL SCHARRE, ARMY OF NONE: AUTONOMOUS WEAPONS AND THE FUTURE OF WAR (W.W. Norton & Co., 2018).

invaluable to the intelligence community and military. Indeed, the National Center for Counterterrorism ["NCTC"] housed in the Office of the Director of National Intelligence ["ODNI"] has a vast repository of data on terrorist networks. In 2011, ODNI contained over half a million names of terrorists.[80] One of its primary mission areas is "identity management," which it defines as the "responsibility to serve as the central and shared knowledge bank on known and suspected terrorist and international terror groups, as well as their goals, strategies, capabilities, and networks of contacts and support."[81] The NCTC gleans this information from a wide variety of sources including classified human intelligence sources, signal intelligence, and open sources like social media.[82] Data about a terrorist network may paint a picture of the connections between terrorists, highlighting the central individuals and showing which persons exist on the fringes of the network.[83] Clearly ". . . data can be . . . analyzed to support the development of either kinetic or non-kinetic strategies to counter terrorism."[84] AI systems could optimize the synthesizing of intelligence, or even predict future terrorist behavior based on the patterns in the data.

Hypothetically, as a counterterrorism prevention and prosecution strategy, a government-controlled AI system could scour social media posts seeking those posts that follow a particular formula: language of threats of violence, declarations of imminence, coupled with some extremist ideological language. If the programmed parameters of the AI system's alerts conform to constitutional principles, this method could benefit law enforcement and prosecutors. However, if law enforcement arrests and prosecutes a person based on evidence gathered using this kind of AI monitoring method, without other evidence, it seems likely that the defendant would challenge the assumptions and parameters used by the AI system, meaning that the AI system would then likely have to be explainable. In other words, the results or conclusions generated by the AI system explained in this way may be able to be understood or justified, not rendered opaque by complex code or by an internal mechanism used to intentionally hide a

---

[80] Roberts & Everton, *supra* note 32, at 2.

[81] NATIONAL COUNTERTERRORISM CTR., TODAY'S NCTC 6, (Aug. 2017), https://www.dni.gov/files/NCTC/documents/features_documents/NCTC-Primer_FINAL.pdf.

[82] "Social media scraping tools" are commercially available and are typically used to "extract data from channels which not only include social networking sites, such as Facebook, Twitter, Instagram, LinkedIn…etc., but also include blogs, wikis, and news sites." *Top 5 Social Media Scraping Tools for 2019*, OCTOPARSE (Oct. 17, 2018), https://www.octoparse.com/blog/top-5-social-media-scraping-tools-for-2018.

[83] *See generally*, Roberts & Everton, *supra* note 32.

[84] *Id.* at 9.

proprietary algorithm.[85] In the context of criminal prosecutions, the importance of understanding the programming and rules in an AI system builds trust in that AI system. If a scientist or programmer cannot evaluate the inner workings of an AI system's algorithm, how can people trust that it is acting in accordance with its programming? By way of example, what if the programmer built a system that purposely targets Muslims only? That kind of unconscionable programming would violate cherished constitutional rights, discriminating against a specific sect of American society. Consequently, this approach would pose substantial prosecutorial risk, as well as undermine bedrock American principles. Indeed, if the government used some kind of algorithmic terrorist risk assessment as a part of sentencing, or to propose potential deradicalization methods for that defendant, such a proposal could potentially violate the defendant's constitutional due process rights.[86] Thus, using AI in the realm of criminal prosecutions seems like a misapplication of the technology – it would be better suited in other contexts.

The U.S. government is limited in how it can tackle the holistic approach to counterterrorism that employs both hard and soft measures. However, the U.S. intelligence community and military already have resources that could benefit from AI optimization. However, much in the same way that the NCTC "crawls" or "scrapes" the Internet by collecting open source information, the U.S. government should use an AI system to "scrape" the social media accounts of foreign terrorists. This use of AI Internet scraping would not only be used for intelligence gathering and non-kinetic measures, but for offensive, cyber-enabled hard measures. The proposal section below will explore some forward-thinking, possibly provocative approaches to AI system deployment, beyond these obvious options already available and likely used in hard counterterrorism approaches.

VI. SOCIAL MEDIA-CREATED ARTIFICIAL INTELLIGENCE AND SOFT COUNTERTERRORISM STRATEGIES

The power and ubiquity of social media platforms like Facebook, Twitter, Instagram, LinkedIn, YouTube, among others,[87] increased

---

[85] KYNDI, HOW 'EXPLAINABILITY' IS DRIVING THE FUTURE OF ARTIFICIAL INTELLIGENCE 2 (2018) https://kyndi.com/wp-content/uploads/2018/01/Kyndi-final-Explainable-AI-White-Paper.pdf.

[86] For a wonderful discussion of the difficulty implementing unexplainable "black box" algorithms in the criminal justice system, *see* Leah Wisser, *Pandora's Algorithmic Black Box: The Challenges of Using Algorithmic Risk Assessments in Sentencing*, 56 AM. CRIM L. REV. 1811 (2019).

[87] Although Google is more of a search engine than a "social media" platform, this discussion

exponentially since they took hold in the early to mid-aughts. They spread like wildfire across the globe and reached *billions* of people. They provide a modality for free expression: individuals create content, share information, communicate with friends, and even make purchases, transfer money, and sell goods or services. As discussed above, the broad reach of social media, and its unique capacity to bring people together creates a place for positive discussion as well as a forum for malefactors to conspire, plan, and recruit for malignant purposes. The uniquely broad reach of these private-sector platforms, in conjunction with special legal status, makes them ideal methods for furthering softer counterterrorism measures.

A. Challenges

Social media platforms walk a fine line between two conflicting goals: the primary goal being to promote a free environment for expression and the subordinate goal being to police objectionable content. Many questions arise about how social media platforms could possibly achieve their goals. The following sections will highlight common challenges to social media platforms' efforts in counterterrorism.

1. § 230 of the Communications Decency Act ["CDA"]

In the nineties, in response to absurd impositions of liability on Internet Service Providers ["ISPs"] based on common law principles, Congress passed the CDA in 1996, but the American Civil Liberties Union challenged the law shortly thereafter, resulting only in one provision remaining intact: § 230.[88] Congress plainly expressed its policy in enacting the CDA, specifically § 230: "(1) to promote the continued development of the internet and other interactive computer services and other interactive media; (2) to preserve the vibrant and competitive free market that presently exists for the internet and other interactive computer services, unfettered by Federal or State regulation. . . ."[89] To achieve its legislative purpose, Congress implemented broad protections for "interactive computer services"[90]: "[n]o provider or user of an interactive computer service shall be

---

captures Google's universal reach into everyday life as well.

[88] Phe, *supra* note 7, at 108. *See also id.*, *supra* note 7, at 103-110 (explaining the impetus for the CDA's implementation and a general historical survey of the CDA).

[89] 47 U.S.C. §§ 230(b)(1) - (b)(2).

[90] Phe, *supra* note 7, at 110. ("Immunity under § 230 extends only to interactive computer services providers and not to information content providers.").

treated as the publisher or speaker of any information provided by another information content provider," and they will enjoy immunity from civil liability for "any action voluntarily taken in good faith to restrict access to or availability of material to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected. . . ."[91] The language of the statute distinguishes between such platform intermediaries and "information content providers," which are "entit[ies] that [are] responsible, in whole or in part for the creation or development of information provided through the Internet or any other interactive computer service."[92] In other words, "information content providers" actually create content and post that content on the Internet, as opposed to acting as a medium or virtual bulletin board for others' content.

The statute's broad grant of immunity from liability under § 230 evolved to apply to social media platforms.[93] The immunity continues to allow a protective shroud around those social media platforms which allow users to post content. Congress granted broad immunity to foster free-speech, free-market, and exchange of ideas over two decades ago. However, when implementing § 230 during the Internet's infancy, Congress could not predict how it would connect billions of people on an individual level. It used to be the major method of connectivity for a few million users, but it has become a foundational, yet subordinate protocol upon which the behemoth social media platforms with billions of users built themselves – platforms like Facebook, Twitter, Google, and YouTube.[94] The effect of § 230 immunity is that social media cannot be held responsible for extremist content and terrorist activity.[95]

---

[91] 47 U.S.C. §§ 230(c)(1)-(2)(A)

[92] 47 U.S.C. § 230(f)(3).

[93] The three-pronged test used by courts to determine whether an entity should receive §230 immunity: "whether an interactive computer service is used or provided; whether the entity can be considered an information content provider of the objectionable content or activity in question; and whether the cause of action seeks to hold the entity as a publisher or speaker of third-party content." Phe, *supra* note 7, at 110.

[94] Phe notes that it is indeed "questionable whether social media websites even qualify for immunity under §230" because "[s]ocial media websites certainly take a more active role in selecting the kinds of information disseminated on their platforms compared to traditional publishers," although courts seem to uniformly apply §230 to social media platforms. *Id.* at 127.

[95] *See e.g.*, Gonzalez v. Google Inc., 282 F.Supp.3d 1150 (N.D. Cal 2017).

2. "Engagement" Problems in Social Media Algorithms

Social media platforms employ illustrious algorithms and AI systems to optimize searches and offer advertisements to third parties about platform users. Indeed, the issue of "engagement," that is, algorithms calibrated to "maximize long-term viewer engagement and satisfaction,"[96] brings about some ironic perversions. The algorithm evidently presents extreme search results related to whatever a person searches for, as opined by former counsel for National Security Agency ["NSA"] and Department of Homeland Security ["DHS"], Stewart Baker:

> [t]he drive to extremes that recommendation engines send you down is very, very real because there's something – and it could be it drives you towards extremes of weightlifting; it drives you towards extremes of home knitting; or it drives you toward extremes of sexual perversion. And part of this is just it's human nature, right? They're telling you what other people who looked at this video went on to look at, and those people went on to look at . . . other extreme things precisely because that was their nature.[97]

The search terms are inputs for the algorithms, and to keep viewers engaged, its design results in outputs of extreme content related to that original term. The user then accesses the content, and by accessing the content, the user inputs his/her new views of the extreme content into the algorithm, creating a self-radicalizing feedback loop.[98] The question then becomes, should social media platforms try to counteract or revamp these engagement models? Baker notes the quandary:

> There is a reinforcing effect. So, I see the problem, I'm not sure the answer is stop doing recommendations because there is an awful lot of value in those recommendations … so figuring out when you should stop or moderate your recommendation engines, strikes me, maybe because I am a minority in this regard, at least in Silicon Valley – they would cheerfully take three quarters of my views and say "we'll never recommend anyone who looks at those" – and so the idea that they should be saying people with views like yours

---

[96] Paige Cooper, *How Does the YouTube Algorithm Work? A Guide to Getting More Views*, HOOTSUITE (Apr. 8, 2019), https://blog.hootsuite.com/how-the-youtube-algorithm-works/.

[97] Stewart Baker, In the *cyber adversary Olympics, it's Russia for the gold and North Korea (!) for the silver*, *THE CYBERLAW PODCAST* (Feb. 27, 2019), https://www.steptoe.com/feed-Cyberlaw.rss.

[98] *Id.*

should not be able to find each other online is troubling.[99]

Along the same vein, a recent article published by the New York Times documents the role played by engagement-oriented algorithms in the radicalization of a young man named Caleb Cain, who found himself in the world of far-right extremism, saying "[w]hen I found this stuff, I felt like I was chasing uncomfortable truths . . . I felt like it was giving me power and respect and authority."[100] Clearly, social media platforms with such engagement models built into their algorithms must weigh the great benefit of those models against the risk of radicalization, which many platforms recently pledged to actively combat. The solution to the problem is certainly not easy to discern.

3. How to Hold Social Media Platforms Accountable

There is no comfortable means of regulating social media platforms, but U.S. policy makers must start finding creative solutions. Some commentators suggest legislation proposing that § 230 "should be revised to categorically withhold immunity from ISPs that have actual knowledge their websites are used to conduct terrorist operations, as well as ISPs that choose to profit off their platforms despite reasonable notice or knowledge that they are hosting content from a terrorist or terrorist organization."[101] Social media platforms should be incentivized to implement baseline self-policing measures for extremist content, or alternatively, disincentivized to allow extremist content. While many platforms declared their commitment to countering extremist behavior, they need something beyond the current framework to make meaningful progress in opposing terrorist activities on their networks.

B. Current Approaches and Strategies for Social Media Use of AI in Counterterrorism

Notwithstanding the general criticism of social media platform efforts to combat terrorist activities, some platforms began implementing AI systems in their self-policing counterterrorism efforts. Facebook's founder and CEO, Mark Zuckerberg, while testifying before Congress, noted several ways that Facebook uses AI to combat objectionable actors on the platform. In the context of foreign disinformation and election meddling, he said that "the AI tools that we deployed in those elections were able to proactively take down

---

[99] *Id.*

[100] Kevin Roose, *The Making of a YouTube Radical*, NEW YORK TIMES, (June 8, 2019), https://www.nytimes.com/interactive/2019/06/08/technology/youtube-radical.html.

[101] Phe, *supra* note 7, at 129.

tens of thousands of fake accounts that may have been trying to [influence those elections]."[102] Beyond election meddling, Facebook targets extremists who virtually gather on its platform: "if there is a group that their primary purpose or a large part of what they do is spreading hate, we will ban them from the platform," and Facebook works to "adjust [its] algorithms to prevent those interested in violence or bad activities from being connected with other like-minded individuals[.]"[103] Facebook's approach certainly interposes barriers for extremists, but that strategy should be but one facet in the holistic, multi-dimensional approach to counterterrorism.

One of the most creative uses of AI systems in counterterrorism is the Jigsaw Redirect Method:

> Jigsaw, the Google-owned tech incubator and think tank . . . has been working . . . to develop a new program it hopes can use a combination of Google's search algorithms and YouTube's video platform to target aspiring ISIS recruits and ultimately dissuade them from joining the group's cult of apocalyptic violence. The program, which Jigsaw calls the Redirect Method . . . places advertising alongside results for any keywords and phrases that Jigsaw had determined people attracted to ISIS commonly searched for. [104]

The Redirect Method identifies specific extremist narratives, finding the content in English and Arabic, and when a person searches for such extremist content, the algorithm uses microtargeting in curated playlists to offer them counter-radicalization content where online ads might be.[105] The project boasts quality results: the project reached 320,906 users, and "redirected" users watched a total of 500,070 minutes of video during its pilot.[106]

CONCLUSIONS & PROPOSALS

The U.S. government and the private sector social media platforms each have roles to play in counterterrorism, but their roles diverge significantly. As the above analysis demonstrates, the government excels at "hard" counterterrorism strategies. It uses

---

[102] *Testimony of Mark Zuckerberg, Facebook: Transparency and Use of Consumer Data*, *Hearing Before the* H. Comm. on Energy and Com., (Apr. 11, 2018) at 46-47 (unofficial transcript).

[103] *Id.* at 45.

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

[105] *See generally A Blueprint for Bypassing Extremism*, THE REDIRECT METHOD, https://redirectmethod.org/downloads/RedirectMethod-FullMethod-PDF.pdf.

[106] *Id.* at 13.

military force, information operations, and psychological operations abroad to combat global terrorism, but also investigates, arrests, and prosecutes suspected terrorists domestically. But, constitutional, legal, and practical issues impede the government's ability to effectively utilize "soft" counterterrorism strategies like counter-messaging and counter-radicalization.

Conversely, social media platforms lack the legal authority or the business incentives to use hard counterterrorism measures because such tools fall squarely in the domain of the government. However, the social media platforms' incentive to self-police (for instance, in order to avoid passage of legislation that could increase potential liability), lack of constitutional constraints, and access to technology make them well suited to pursue softer counterterrorism strategies.

Consequently, a counterterrorism partnership between the government and social media platforms would provide a comprehensive approach to counterterrorism that marries hard measures executed by the government, with soft strategies implemented by social media platforms.

With these conclusions in mind, this analysis will propose useful ways the government and social media platforms can deploy AI systems to augment their respective strengths to combat terrorist behavior.

1. The United States Government Should Deploy AI Systems to Coordinate Offensive Cyber-Capabilities Against Foreign Terrorist Recruiters on Social Media

The U.S. military has express authority from Congress in the John S. McCain National Defense Authorization Act for Fiscal Year 2019 ["NDAA FY2019"] to use cyber-capabilities to battle cyber-threats: "[i]t shall be the policy of the United States, with respect to matters pertaining to cyberspace, cybersecurity, and cyber warfare, that the United States should employ all instruments of national power, including the use of offensive cyber capabilities, to deter if possible, and to respond when necessary, all cyber[-]attacks or other malicious cyber activities of foreign powers that target United States interests . . . ."[107] The term "foreign powers" includes groups "engaged in international terrorism or activities in preparation therefor,"[108] and ISIS

[107] National Defense Authorization Act Fiscal Year 2019, H.R. 5515, 115th Cong., Title XVI § 1636(a) (2018).

[108] Foreign Intelligence Surveillance Act ["FISA"], 50 U.S.C. § 1801(a)(4) (2012). The NDAA FY 2019 adopts this definition of "foreign power" in its authorization of offensive cyber-capabilities by the government.

recruitment or encouragement on social media likely counts as "preparation" for international terrorism if done by a non-U.S. person. While the U.S. government would be limited in using its capabilities against U.S. persons abroad,[109] such limitations do not apply to non-U.S. persons abroad. Although, at least for surveillance of non-U.S. persons, the law requires minimization procedures be implemented to prevent accidental or incidental surveillance of U.S. persons.[110] Moreover, U.S. constitutional rights do not apply to non-U.S. persons located abroad, as held by the Supreme Court in *Verdugo-Urquidez*.[111]

The advantages of machine learning AI systems could be directly implemented in an aggressive, cyber-enabled information operation on social media platforms. An AI controlled by the U.S. government could be trained to crawl social media platforms like Facebook or YouTube using a false identity "cover account," designed to appear like a person susceptible to recruitment. The AI system would be trained on terrorist recruitment data (e.g., name, location, interests, language in posts, etc.), which would allow it to generate an output with all the characteristics exemplary of a person sought for recruitment by a terrorist organization like ISIS. The AI system would be able to create many of these false identity accounts.

The AI would be programmed to generate apparently extremist content by crawling extremist sites or archived extremist data, training on language used by extremists.[112] The AI system could then peruse or periodically post extremist-leaning comments on extremist pages, indicative of a person who appears ready to be radicalized, with the goal of luring a terrorist recruiter into making contact. Once the recruiter makes contact, the AI system would engage in dialogue with the recruiter, until the AI system decides it has an opportunity to strike. The AI system would then offer a hyperlink to the recruiter and the hyperlink would purport to activate a video chat (like Skype), encrypted messenger (like telegram), or show an extremist website or picture. That link, however, would be a vehicle for a "Network Investigative Technique[] ["NIT"] . . . which include computer code that investigators can send covertly to a device," which "can send law enforcement particular information, often including the device's true

---

[109] *See e.g.*, FISA, 50 U.S.C. §§ 1881(b)-1881(c) (2012).

[110] *See* FISA, 50 U.S.C. § 1881(a) (2012) (requiring minimization procedures for foreign surveillance of non-U.S. persons to ensure no incidental surveillance of U.S. persons).

[111] Verdugo-Urquidez, *supra* note 69, at 265.

[112] *See e.g.*, Nicole Martin, *Did A Robot Write This? How AI Is Impacting Journalism*, FORBES (Feb. 8, 2019), https://www.forbes.com/sites/nicolemartin1/2019/02/08/did-a-robot-write-this-how-ai-is-impacting-journalism/#2b84b9097795 (suggesting AI is already capable of "journalism").

IP address – which investigators can use to identify the subscriber and user of the device."[113] More sophisticated malware can allow the operators to infiltrate the system without the operator's knowledge. Being able to access a terrorist recruiter's computer would potentially offer troves of intelligence, could shut down a recruiter's computer or operation, and/or infiltrate the recruiter's phone or computer and copy his/her information.[114] The ability of AI systems to synthesize vast amounts of data about terrorist recruiters, generate responses in social media, while coordinating between accounts, would give an AI system an edge over the numerous human operators it would take to orchestrate such a campaign.

The above proposed cyber-campaign against terrorist recruiters assumes that the targets are non-U.S. persons because U.S. persons are entitled to the panoply of rights guaranteed under the Constitution, even when located abroad. How can U.S. government cyber-operators be sure that a person is not a U.S. person? AI systems can potentially solve that problem with their incredible data processing abilities. An AI system can likely scrape location data from accounts in the same way that advertisers obtain such data. If properly programmed to ignore people linked to the United States, the AI system can avoid catching U.S. persons in its operational net. Moreover, any statutory minimization procedures required could presumably be programmed into the AI systems engaged in cyber-operations.

AI systems may be the perfect tool for hard counterterrorism measures on social media. If the government tailored AI systems to act like susceptible potential recruits, they could coordinate automated, well-orchestrated cyber-campaigns to dismantle terrorist recruitment networks abroad, while devoting less human capital, and not violating upon other States' sovereignty.

2. Social Media Platforms Should Use AI Systems to Inundate Extremist Fora with Counter-Messaging

Social media platforms should embrace their exclusion from constitutional requirements to push harder against terrorist recruitment on their sites, a problem for which AI offers a solution.

As discussed *supra*, the Saudi government's anti-/counter-radicalization program *Sakina* aims to have Muslim scholars enter

---

[113] U.S. DEPARTMENT OF JUSTICE, REPORT OF THE ATTORNEY GENERAL'S CYBER DIGITAL TASKFORCE 53 (July 2, 2018), https://www.justice.gov/ag/page/file/1076696/download.

[114] *See e.g.*, Ben Buchanan, *What to Make of Cyber Command's Operation Against the Internet Research Agency*, LAWFARE (Feb. 28, 2019), https://www.lawfareblog.com/what-make-cyber-commands-operation-against-internet-research-agency.

extremist social media fora to challenge radical views and engage in dialogue for the purpose of dissuading young people susceptible to radicalization in seeking answers in those fora. The Saudi Government *Sakina* program boasted a relatively high success rate.

Social media platforms, using machine learning AI systems, could train AI systems on fundamental principles of Islam taken from non-extremist viewpoints and essentially implement a worldwide "AI based-*Sakina*" program that would use AI-created accounts to debate extremists in extremist fora. AI systems could generate account handles, and after training on vast amounts of data, could engage extremists or individuals in the midst of radicalization, while challenging the extremist narratives with difficult philosophical questions modeled on extent community beliefs or factual statements about the extremist movement's assumptions. AI-powered "chatbots" have been tested in clinical psychological settings and in some cases, successfully used to replace one-on-one therapy sessions.[115] The perspectives offered could draw from wide ranging viewpoints, so long as they exclude the radical ones. In addition, this strategy could be used in the context of ISIS-type extremist to domestic white nationalist recruitment. In an ISIS context, challenging the "utopia" narrative might dispel the notion that the ISIS caliphate is a paradise with beautiful land, peace, and jobs. The AI system could, in a manner akin to the Redirect Method, offer access to relevant content like pictures, articles, and videos that challenge those ideas. This strategy, if employed by the government, would implicate the Establishment Clause, but social media platforms, as private actors, do not confront such constitutional barriers.

The Redirect Method is the most creative soft approach employed by social media, but it tackles the issue in a passive fashion. In a realm where the U.S. government cannot tread easily, a *Sakina* style program deployed by social media platforms would allow aggressive infiltration of extremist fora, and affirmatively and offensively present counter-radicalization material to people it deems to be on the verge of recruitment.

To achieve a holistic approach to terrorism that implements both hard and soft measures, the U.S. government and the social media platforms must attack the problem using their strengths, and that requires a concomitant and coordinated effort involving the use of AI.

---

[115] Kayla Matthews, *Therapy Chatbots are Transforming Psychology*, MEDIUM (Apr. 10, 2018), https://chatbotslife.com/therapy-chatbots-are-transforming-psychology-de67570236bc.