ARTIFICIAL INTELLIGENCE BASED SUICIDE PREDICTION

Mason Marks*

ABSTRACT

Suicidal thoughts and behaviors are an international public health concern contributing to 800,000 annual deaths and up to 25 million nonfatal suicide attempts. In the United States, suicide rates have increased steadily for two decades reaching 47,000 per year and surpassing annual motor vehicle deaths. This trend has prompted government agencies, healthcare systems, and multinational corporations to invest in tools that use artificial intelligence to predict and prevent suicide. This article is the first to describe the full landscape of these tools, the laws that apply to their operation, and the underexplored risks they pose to patients and consumers.

AI-based suicide prediction is developing along two separate tracks: In "medical suicide prediction," AI analyzes data from patient medical records; In "social suicide prediction," AI analyzes consumer behavior derived from social media, smartphone apps, and the Internet of Things. Because medical suicide prediction occurs within the healthcare system, it is governed by laws such as the Health Information Portability and Accountability Act (HIPAA), which protects patient privacy; regulations such as the Federal Common Rule, which protects the safety of human research subjects; and general principles of medical ethics such as autonomy, beneficence, and justice. Moreover, medical suicide prediction methods are published in peer-reviewed academic journals. In contrast, social suicide prediction typically occurs outside the healthcare system where it is almost completely unregulated, and corporations often maintain their prediction methods as proprietary trade secrets. Due to this lack of transparency, little is known about their safety or effectiveness. Nevertheless, unlike medical suicide prediction, which is primarily experimental, social suicide prediction is deployed globally to affect people's lives every day.

Though AI-based suicide prediction may improve our understanding of suicide while potentially saving lives, it raises many risks that have been underexplored. The risks include stigmatization of people with mental illness, the transfer of sensitive health data to third-parties such as advertisers and data brokers, unnecessary involuntary confinement, violent confrontations with police, exacerbation of mental health conditions, and paradoxical increases in suicide risk. After describing these risks, the article presents a policy framework for promoting safe, effective, and

*Research Scholar, NYU Law School Information Law Institute; Visiting Fellow, Yale Law School Information Society Project, External Doctoral Candidate, Leiden Law School Center for Law and Digital Technologies. JD, Vanderbilt University Law School; MD, Tufts University School of Medicine; BA, Amherst College. Thank you to Katherine Strandburg, Ann Bartow, Ari Waldman, Andrea Matwyshyn, and Roger Ford for their comments on an earlier draft of this article. Thank you to Jack Balkin, Abbe Gluck, Katherine Kraschel, Joel Reidenberg, Adam Pan, Phil Yao, the Yale Solomon Center for Health Law & Policy, the Yale Information Society Project, the Center on Law and Information Policy at Fordham Law School, and the Innovation Center for Law and Technology at New York Law School.

fair AI-based suicide predictions. The framework could be adopted voluntarily by companies that make suicide predictions or serve as a foundation for regulation in the US and abroad.

TABLE OF CONTENTS

INTRODUCTION

Suicide is a global public health concern. The World Health Organization (WHO) estimates it claims a life every 40 seconds and kills 800,000 per year.[1] Non-fatal suicide attempts may be 20 to 25 times more common. Both attempted and completed suicides take a large toll on families, communities, and healthcare systems, and they are on the rise.[2] In the US, suicide rates rose by 25% between 1999 and 2016, and half the states experienced a rise of over 30%.[3] Suicide is now the second leading cause of death in American teens, it kills more Americans each year than auto accidents or homicides, and it costs the US economy over $69 billion dollars a year.[4]

To address rising suicide rates, governments, healthcare systems, and corporations are developing artificial intelligence (AI) based suicide prediction tools. In theory, suicide can be prevented if it can be accurately predicted. Yet in practice, predicting suicide is challenging because it is a complex problem with many contributing factors. Traditional methods of prediction involve questionnaires that yield inaccurate results; often little more accurate than a coin toss, or what would be expected due to chance.[5] AI shows promise for increasing the accuracy of these predictions.[6]

This article is the first to describe the full range of AI-based suicide prediction tools and their legal, ethical, and public health implications. Healthcare systems including Vanderbilt University Medical Center, government agencies such as the US Department of Veterans Affairs (VA), and private companies including Facebook are developing AI-based suicide prediction tools.[7] Though these tools have the potential to identify people at high risk for suicide permitting intervention and possibly prevention, they also potentially violate people's privacy, marginalize vulnerable populations, stigmatize and traumatize people with disabilities, inaccurately

---

[1] World Health Organization, Suicide Data, http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/ (last visited Sep. 25, 2018).

[2] American Foundation for Suicide Prevention, Suicide Statistics, https://afsp.org/about-suicide/suicide-statistics/ (last visited Oct. 14, 2018); World Health Organization, Preventing Suicide – A Global Imperative, http://apps.who.int/iris/bitstream/handle/10665/131056/9789241564779_eng.pdf;jsessionid=A1C3BA3BB3E15829DD187BD773E9A0CF?sequence=1 (last accessed Oct. 14, 2018)

[3] Centers for Disease Control and Prevention, *Suicide rising across the US,* https://www.cdc.gov/vitalsigns/suicide/ (last visited Aug. 21, 2018) (reporting that since 1999, half the U.S. states experienced an increase in suicide rates of over 30%); Sabrina Tavernise, *U.S. Suicide Rate Surges to a 30-Year High,* NY TIMES (Apr. 22, 2016), https://www.nytimes.com/2016/04/22/health/us-suicide-rate-surges-to-a-30-year-high.html

[4] Alicia Vanorman and Beth Jarosz, *Suicide Replaces Homicide as Second-Leading Cause of Death Among U.S. Teenagers,* Population Reference Bureau (Jun. 9, 2016), https://www.prb.org/suicide-replaces-homicide-second-leading-cause-death-among-us-teens/; National Highway Traffic Safety Administration, *USDOT Releases 2016 Fatal Traffic Crash Data,* https://www.nhtsa.gov/press-releases/usdot-releases-2016-fatal-traffic-crash-data (reporting that in 2016, nearly 45,000 American died by suicide. By comparison, 37,461 Americans we killed in auto accidents); Federal Bureau of Investigation, *Murder, https://ucr.fbi.gov/crime-in-the-u.s/2016/crime-in-the-u.s.-2016/topic-pages/murder* (last visited Sep. 25, 2018) (reporting that in 2016, there were 17,250 U.S. homicides); Margot Sanger-Katz, *Gun Deaths Are Mostly Suicides,* NY TIMES, (Oct. 8, 2015) https://www.nytimes.com/2015/10/09/upshot/gun-deaths-are-mostly-suicides.html

[5] Colin G. Walsh et al., *Predicting Risk of Suicide Attempts Over Time Through Machine Learning,* 5 CLINICAL PSYCHOLOGICAL SCI. 1, 2 (2017).

[6] *Id.*

[7] Martin Kaste, *Facebook Increasingly Reliant on A.I. To Predict Suicide Risk,* ALL THINGS CONSIDERED (Nov. 17, 2018), https://www.npr.org/2018/11/17/668408122/facebook-increasingly-reliant-on-a-i-to-predict-suicide-risk.

categorize people as suicidal or non-suicidal, promote unnecessary hospitalization and forced medication (and in some parts of the world incarceration), exacerbate mental health conditions, and paradoxically increase the risk of suicide. These risks have received little or no attention in the media and academic literature. As AI-based suicide prediction tools become more widespread, it is important to determine whether they are helping to prevent mental illness and suicide or contributing to these problems.

The article consists of three parts. Part I explains why traditional methods of suicide prediction inaccurate and how AI-based tools may improve upon their accuracy. These tools fall into two general categories: "medical" and "social" suicide prediction, which use different methods and draw from different data sets. Part II describes how these two categories are governed by different laws leaving medical suicide prediction heavily regulated and social suicide prediction almost completely unregulated.

Part II describes the individual and societal risks of AI-based suicide prediction and how they may disproportionately impact vulnerable populations. The risks are divided into privacy, safety, and autonomy harms. Part II also explains how suicide prediction is analogous to predictive policing and suffers from similar shortcomings and misconceptions. This analogy is fitting because in some countries, attempted suicide is a criminal offense, and AI-based suicide prediction could result in criminal penalties including fines and imprisonment.

Part III presents a policy framework for more responsible implementation of medical and social suicide prediction, which is designed to maximize safety and effectiveness while minimizing the risk of harm to patients and consumers. Healthcare providers and companies that make suicide predictions could voluntarily adopt the framework in the form of industry standards, or the framework can serve as a template for drafting laws to regulate suicide predictions in the US and internationally.

I. AI May Improve the Accuracy of Suicide Predictions

A. *Traditional Methods of Suicide Prediction Are Inaccurate*

Traditionally, doctors and therapists predicted suicide by administering written questionnaires to patients. The answers were converted into scales thought to reflect suicide risk. Typical examples include the Suicide Intent Scale (SIS), the Scale for Suicidal Ideation (SSI), and the Beck Hopelessness Scale (BHS). However, their predictive abilities are unimpressive: "Recent meta-analyses of hundreds of studies from the past 50 years indicate that the ability to predict future suicide attempts has always been at near chance levels."[8] According to one large study: "All of the scales and tools reviewed here had poor predictive value."[9]

[8] Walsh *supra* note 5.
[9] MKY Chan et al., *Predicting suicide following self-harm: systematic review of risk factors and risk scales*, 209 British J. Psychiatry 277, 279 (2016).

Suicide is difficult to predict because it is a complex problem with many risks and contributing factors.[10] There is no single risk factor that reliably predicts self-harm. Though there is a clear association between suicide attempts and some variables such as depression and substance use disorders, most people with these conditions do not attempt suicide.[11] Other risk factors include anxiety disorders, bipolar disorder, eating disorders, unemployment, a family history of suicide, having been released recently from a psychiatric hospital, "belonging to a sexual minority" group, "infection with the brain-tropic parasite *Toxoplasma gondii*, and "childhood physical, sexual, or emotional abuse."[12] Because these risk factors are so numerous and diverse, it is difficult to account for them all in a single predictive model.

Suicide prediction is also hindered by the fact that suicide is relatively rare.[13] Though on a national and global scale, the number of people who die by suicide is not trivial, only a very small percentage of people under psychiatric care attempt suicide.[14] Complicating matters, while suicide is relatively uncommon, its risk factors, such as suicidal thoughts, are extremely common.[15] According to estimates by the US Substance Abuse and Mental Health Service Administration, 9.8 million American adults seriously contemplated suicide in 2015.[16] However, only 2.7 million formulated concrete suicide plans and about 1.4 million made suicide attempts.[17] These statistics demonstrate that even though suicidal thoughts are relatively common, and they are a risk factor for suicide, most people who have suicidal thoughts do not attempt suicide.[18] The same can be said for other risk factors such as major depressive disorder, which is estimated to affect over 16 million American adults.[19] When the frequency of a medical condition is low, and that of its risk factors is high, the predictive ability of tests for that condition may also be low and may produce many false positives.[20]

Suicide prediction is also challenging because talking about suicide is taboo. People with suicidal thoughts may be afraid to discuss them with friends, family, and healthcare providers out of fear they might be judged, stigmatized, or hospitalized and medicated against their will.[21] In fact, most people who commit suicide (about 70%) never disclose their suicidal thoughts to physicians.[22] Certain subpopulations may share cultural values that make discussion of suicidal

---

[10] Gustavo Turecki and Brent A. David, *Suicide and Suicidal Behavior*, 387 LANCET 1227 (2016) (reporting genetic, developmental, and social risk factors for suicide).
[11] Citation for most people with depression or substance use disorders don't attempt suicide
[12] Turecki *supra* note 10.
[13] *See* Steffan Davies et al., *Depression, suicide, and the national service framework – Suicide is rare and the only worthwhile strategy is to target people at high risk*, 322 BMJ 1500 (2001).
[14] Roger Mulder et al., *The Futility of Risk Prediction in Psychiatry*, 209 BRITISH J. PSYCHIATRY 271, (2016).
[15] https://www.youtube.com/watch?v=6QPI0AFQ_Yw
[16] *9.8 Million American adults had serious thoughts of suicide in 2015*, SUBSTANCE ABUSE MENTAL HEALTH SERVICES ADMIN. (Sep. 15, 2016), https://www.samhsa.gov/newsroom/press-announcements/201609150100.
[17] *Id.*
[18] *See Id; See* also Chris Poulin et al., *Predicting the Risk of Suicide by Analyzing the Text of Clinical Notes*, 9 PLOS ONE e85733 (2014) (reporting that most people who have suicidal thoughts do not attempt suicide).
[19] *Major Depression*, NAT'L INST. MENTAL HEALTH, https://www.nimh.nih.gov/health/statistics/major-depression.shtml#part_155028 (last visited Jan. 2019).
[20] *See* Karlijn J. van Stralen et al., *Diagnostic methods I: sensitivity, specificity, and other measures of accuracy*, 75 KIDNEY INT'L 1257, 1261 (2009).
[21] Lindsay Sheehan et al., *The specificity of public stigma: A comparison of suicide and depression-related stigma*, 256 PSYCHIATRY RESEARCH 40 (2017).
[22] Most people who commit suicide never disclose thoughts or plans to their doctor

thoughts more challenging. For example, the US military's culture of promoting mental toughness, self-sacrifice, and control and suppression of emotions can serve as an obstacle to frank discussion of emotionally charged issues like suicide.[23] Service members may feel obligated to suppress their feelings and "shake it off" when facing feelings of despair.[24] The following section explains how AI may increase our ability to identify people at risk for suicide and describes the two paths of AI-based suicide prediction.

B. *The Two Tracks of AI-based Suicide Prediction*

AI may overcome many limitations of traditional suicide screening tools and increase the accuracy of predictions. AI-based prediction tools can be divided into two broad categories: The first involves analysis of patient medical records. It is performed by doctors, public health researchers, government agencies, hospitals, and healthcare systems. I refer to this category as "medical suicide prediction" because it is based on medical records and is usually conducted within the healthcare system; The second category involves analysis of consumer behavior and social interaction derived from retail purchases, smart phone apps, social media, and other commercial activities outside of healthcare. I refer to this category as "social suicide prediction" because it is based on data derived from people's technology-mediated interactions and transactions.

In the US, medical and social suicide prediction are subject to different laws. Medical suicide prediction is governed by the Health Information Portability and Accountability Act (HIPAA), which protects patient privacy and imposes civil and criminal penalties on covered entities when patient records are breached.[25] It is subject to the Federal Common Rule, which safeguards human research subjects and requires research protocols to be approved by hospital and university institutional review boards (IRBs).[26] All research must also comply with general principles of medical ethics, such as autonomy, respect for persons, beneficence, and justice.[27] These laws, regulations, and principles protect the privacy, safety, and autonomy of people subjected to medical suicide prediction. In contrast, these laws and standards typically do not apply to social suicide prediction because it occurs outside the healthcare system. Instead, because it involves predictions about consumers, it is governed by agencies that protect consumers and regulate interstate commerce and communication such as the Federal Trade Commission (FTC), the Federal Communications Commission (FCC), and potentially the Food and Drug Administration (FDA). At least so far, these agencies have taken little or no interest in suicide predictions and their associated risks.

There are some exceptions to the above observations. A few groups conduct both medical and social suicide predictions. For instance, the VA has analyzed veterans' medical records and their

[23] Chris Poulin et al., *Predicting military and veteran suicide risk: cultural aspects*, 9 PLOS ONE 1 (2014).
[24] *Id.*
[25] 45 C.F.R. §160.301 (2000).
[26] 45 C.F.R. §46.109 (2018).
[27] *See Id.; see* also National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research* (1979).

social media activity.[28] When the VA makes social suicide predictions, they are subject to laws that typically cover medical suicide predictions because the VA is a covered entity under HIPAA and it must comply with the Common Rule. Similarly, tech companies such as Amazon and Microsoft, which typically market their products to consumers, are increasingly investing in the healthcare industry. If their products utilize medical records, then they must comply with laws that typically govern medical suicide prediction. However, despite these exceptions, for the most part, groups usually conduct either medical or social suicide predictions, and they can typically be divided into groups that reside within the healthcare system ("medical suicide predictors") and those that do not ("social suicide predictors"). Though suicide predictors can be sorted into these groups, there is considerable variation within each category with respect to the populations studied, the data collected, and the methods used. The following sections describe the activities of the most prominent medical and social suicide predictors.

1. Medical Suicide Prediction

Medical suicide prediction uses AI to scan and analyze medical records. It is most often performed by academic medical centers, hospitals, and government agencies such as the VA. It can be further subdivided into experimental and clinical suicide prediction. Experimental suicide prediction is performed for research only and is usually done retrospectively by analyzing the records of patients who have already attempted or completed suicide. In contrast, clinical suicide prediction is performed prospectively to predict suicide in living patients, and it is used to inform future treatment decisions for those patients. For the time being, medical suicide prediction is primarily experimental, and except for one program at the VA, it is not widely used to guide clinical decision-making.

a. Hospitals and Healthcare Systems

In one of the largest studies to date, Simon et al. analyzed the records of nearly 3 million patients across seven health systems in multiple states.[29] IRBs representing each system approved the study design.[30] The records included data from over 10 million mental health specialty visits and nearly 10 million primary care visits.[31] The variables studied included "313 demographic and clinical characteristics" including existing medical, mental health, and substance use diagnoses, medications, and commonly administered depression questionnaires.[32] The authors report that their method of combining large volumes of medical records with data from standard mental health questionnaires outperformed previous methods using medical records alone.[33] According to Simon: "people with risk scores in the highest five percent accounted for almost half of

[28] Chris Poulin and Gregory Peterson, *Mobile and social networking technology monitors big data from messages to detect suicide risk in military veterans*, ELSEVIER (Nov. 11, 2015), https://www.elsevier.com/connect/artificial-intelligence-app-combats-suicide-in-veterans.
[29] Gregory E. Simon et al., *Predicting Suicide Attempts and Suicide Deaths Following Outpatient Visits Using Electronic Health Records,* 175 AMER. J. PSYCHIATRY 951, 953 (2018).
[30] *Id.*
[31] *Id.*
[32] *Id.* at 953.
[33] *Id.* at 958.

suicide attempts compared to about one third with previous models."[34] However, despite these impressive results, it is important to point out the limitations of this predictive model. Though it may be an improvement over previous methods, it still produces many false positives and false negatives.[35] Half the people who committed suicide were calculated to be outside the top five percent for suicide risk. Nevertheless, Simon reports his prediction tools are accurate enough to help doctors identify patients at high risk and watch for additional warning signs such as missed appointments.[36]

A separate study of patient data from a single healthcare system was published by Colin Walsh et al.[37] It analyzed the records of 5,167 adults treated at Vanderbilt University Medical Center.[38] The study protocol was approved by the Center's IRB.[39] Unlike Simon's research, Walsh's study focused on people with prior histories of self-injury.[40] 3,250 people were judged by experts to have histories of non-lethal suicide attempts.[41] This group was compared to 12,695 patients from the same Vanderbilt database with no documented suicide attempts.[42] The authors report the accuracy of their suicide prediction models in terms of accuracy under the curve (AUC) where an AUC of 0.5 represents "accuracy no better than chance" and an AUC of 1.0 represents perfect accuracy.[43] Remember that traditional methods of suicide prediction may be little more accurate than a coin flip (a probability of about 50% or 0.50). For patients attempting suicide for the first time, Walsh reported AUC values ranging from 0.82 "at 7 days prior to suicide attempts" to 0.75 "at 720 days prior to suicide attempts."[44] The AUC values decreased relatively linearly as the time before suicide attempts increased.[45] In other words, "model performance steadily improved as the suicide attempt became more imminent."[46]

A smaller study published by Poulin et al. analyzed the clinical records of 100 veterans who died by suicide in 2009.[47] Data from this group was compared to that of a second group with no history of treatment for mental illness and a third group who had previously been hospitalized for psychiatric reasons but did not complete suicide.[48] The study identified words and word pairs in clinical notes that were associated with suicide.[49] Predictive models based on single words, such as "agitation" and "analgesia," had an average predictive accuracy of 59%.[50] The predictive

[34] National Institute of Mental Health, *Predicting Suicide Attempts and Suicide Deaths Using Electronic Health Records – Now Model Substantially Outperforms Existing Suicide Risk Tools,* SCI. UPDATE (Jul. 12, 2018), https://www.nimh.nih.gov/news/science-news/2018/predicting-suicide-attempts-and-suicide-deaths-using-electronic-health-records.shtml.
[35] *Id.*
[36] *Id.*
[37] Walsh *supra* note 5.
[38] *Id.*
[39] *Id.* at 4.
[40] *Id.* at 1.
[41] *Id.*
[42] *Id.* at 4.
[43] *Id.* at 3.
[44] *Id.* at 7.
[45] *Id.* at 8.
[46] *Id.*
[47] Poulin *supra* note 23 at 2.
[48] *Id.*
[49] *Id.* at 3.
[50] *Id.*

accuracy of word pairs ranged from 52% - 69%.[51] Interestingly, models based on word pairs demonstrated higher mean predictive accuracy (64%) than those based on single words (59%), word triplets (60%), and phrases (62%).[52] The study design was approved by the Dartmouth College Center for the Protection of Human Subjects and two VA-affiliated ethics review boards, all of which waived the requirement for obtaining informed consent.[53] This medical suicide prediction research was used as the foundation for a social suicide prediction program, called the Durkheim Project, which is described further in the section on social suicide prediction.

The studies by Simon, Walsh, and Poulin illustrate the heterogeneity of medical suicide prediction programs; each study involved different populations and used different methods. Nevertheless, they share many common features: the studies were conducted by highly-trained physician researchers who are experts in their fields, their protocols were approved by IRBs, they were performed retrospectively and for research only, and their results were published in peer reviewed scientific journals.

b. U.S. Department of Veteran's Affairs REACH VET Program

Unlike the studies described above, which were conducted for research only, the VA performs clinical suicide prediction though a program called Recovery Engagement and Coordination for Health - Veterans Enhanced Treatment (REACH VET). The program uses AI to scan veterans' medical records and identify those at high risk for suicide.[54] It was piloted in a handful of VA hospitals in late 2016 and implemented nationwide in early 2017. The predictive model is refreshed once per month, and following each update, healthcare providers receive notifications regarding patients ranked in the top 0.1 percent for suicide risk.[55] Eventually, it will be expanded to include veterans at more moderate risk (e.g. those ranked in the top 5 percent for suicide risk).[56]

REACH VET incorporates data such as whether veterans take medication for chronic pain, insomnia, or mental illness and how many times they have visited emergency rooms.[57] In the future, it will include additional data such as each veteran's zip code and local unemployment rate.[58] Unlike the algorithms from the experimental studies described above, REACH VET's predictions are used to make clinical decisions. When providers receive notification that veterans are at high risk for suicide, they usually contact the veterans by phone to inquire whether they might benefit from additional attention and support.[59] This intervention can be thought of as a

---

[51] *Id.*
[52] *Id.* at 4.
[53] *Id.* at 5.
[54] Quil Lawrence, *VA Studying Suicide Prevention in Veterans*, NPR MORNING EDITION (Sep. 27, 2017), https://www.npr.org/2017/09/27/553917919/va-studying-suicide-prevention-in-veterans.
[55] Defense Suicide Prevention Office, REACH VET – Predictive Analytics for Suicide Prevention, http://www.dspo.mil/Portals/113/Documents/2017%20Conference/Presentations/REACH%20VET%20Predictive%20Modeling.pdf?ver=2017-08-10-132615-843 (last visited Oct. 5, 2018).
[56] *Id.*
[57] *Id.*
[58] *Id.*
[59] Lawrence *supra* note 41; Jeanette Steele, *Can Math Solve the VA's 20-a-day Suicide Problem?*, SAN DIEGO UNION TRIB. (Apr. 28, 2017), http://www.sandiegouniontribune.com/military/veterans/sd-me-suicide-prediction-20170426-story.html.

"soft-touch" approach because it respects patient autonomy leaving patients in control of what happens next. Below, in the section on social suicide prediction, soft-touch interventions will be contrasted with "firm-hand" interventions that are more severe and may be imposed on people against their will.

Early results from REACH VET have been described as "promising but not definitive" by Dr. Sarah Landes, a clinical psychologist who leads an ongoing clinical trial to evaluate the program.[60] According to Landes: "Early reports from the field demonstrated positive feedback from veterans," and "We have initial data to support that the program is acceptable and feasible."[61]

c. Corporate Medical Suicide Prediction

The previous sections described medical suicide prediction programs implemented by hospitals and healthcare systems. But US healthcare is undergoing rapid technological change. The world's largest tech companies, such as Google, Amazon, Apple, and Microsoft, are investing heavily in the healthcare sector through partnerships with hospitals, pharmacies, and healthcare providers. In some cases, they are storing and analyzing electronic medical records to infer diagnoses and predict clinical outcomes.

i. Google Brain

In 2018, Google scientists published a study describing their use of AI to analyze thousands of electronic health records (HER) to predict clinical outcomes.[62] They validated their approach "using de-identified EHR data from two academic medical centers with 216,221 adult patients hospitalized for at least 24 hours."[63] Their analysis yielded "46,864,534,945 data points, including clinical notes."[64] Deep learning models, a type of AI, achieved high accuracy for tasks such as predicting in-hospital mortality and diagnoses at the time of discharge. According to the study, these models outperformed traditional, clinically-used predictive models in all cases.[65] Though the project was not designed to predict suicide, it shows that Google is experimenting with AI-based methods for making diagnoses and predicting mortality. Mental health diagnoses and medical suicide prediction are likely within its capabilities.

---

[60] Mike Richman, *Study evaluates VA program that identifies vets at highest risk for suicide,* U.S. DEPT. VETERANS AFFAIRS (Sep. 20, 2018), https://www.research.va.gov/currents/0918-Study-evaluates-VA-program-that-identifies-Vets-at-highest-risk-for-suicide.cfm.
[61] *Id.*
[62] Alvin Rajkomar et al, *Scalable and accurate deep learning with electronic health records,* 1 NPJ DIGITAL MED. 1 (2018).
[63] *Id.*
[64] *Id.*
[65] *Id.* at 3.

i. Amazon Comprehend Medical

In 2018, Amazon began marketing software called Amazon Comprehend Medical, which uses AI to identify and analyze text-based medical information.[66] Specifically, the software can extract health data, such diagnoses and medications, from text files and identify relationships between data points.[67] Thus, it can likely be adapted for making suicide predictions.[68] In its current form, the software might predict whether a patient will develop depression, and that information could be conveyed to insurers or other third parties who might use it to market their products.[69] These privacy risks will be discussed further in part II.

2. Social Suicide Prediction

The medical suicide prediction programs describe above are designed to analyze medical records. They can be contrasted with the social suicide prediction programs of companies such as Facebook, Crisis Text Line, and Operation Zero described below. Unlike healthcare providers and medical researchers, these companies lack access to patient records. Instead, they have access to large data sets derived from the behavior of their users. When consumers browse the Internet, shop online, stream music and video, or post on social media, they leave behind trails of digital traces that reflects where they have been and what they have done. Companies collect these digital traces and analyze them with AI to infer people's health information.[70] The goal of social suicide prediction is to calculate suicide risk from those digital traces.

a. Facebook

Social media platforms have an interest in locating and removing suicide-related content. Since Facebook introduced its live-streaming service "Facebook Live" in early 2016, dozens of users have broadcast suicide attempts in real-time on the platform.[71] On February 16, 2017, Facebook CEO Mark Zuckerberg announced the company was developing AI to analyze and flag user-generated content for review by its community managers.[72] In this announcement, Zuckerberg mentioned suicide prediction and prevention as one of his priorities. On March 1, 2017,

---

[66] Melanie Evans and Laura Stevens, *Big Tech Expands Footprint in Health*, WALL ST. J. (Nov. 27, 2018), https://www.wsj.com/articles/amazon-starts-selling-software-to-mine-patient-health-records-1543352136.
[67] Amazon Comprehend FAQs, https://aws.amazon.com/comprehend/faqs/
[68] Knowledge@Wharton, *Amazon, AI and Medical Records: Do the Benefits Outweigh the Risk?* WHARTON SCH. (Dec. 7, 2018), http://knowledge.wharton.upenn.edu/article/amazon-medical-records/.
[69] *Id.*
[70] *See* Marks *supra* note.
[71] *See* Nicolas Vega, *Facebook: We can't stop all live-stream suicides*, NY POST, (Oct. 25, 2017) https://nypost.com/2017/10/25/facebook-we-cant-stop-all-live-stream-suicides/; Jessica Guynn, *Facebook Live is scene of another suicide' police say 'I hope this isn't a trend,'* USA TODAY, (Apr. 26, 2017) https://www.usatoday.com/story/tech/news/2017/04/26/facebook-live-another-suicide/100941914/.
[72] Diana Kwon, *Can Facebook's Machine-Learning Algorithms Accurately Predict Suicide*, SCI. AM. (Mar. 8, 2017), https://www.scientificamerican.com/article/can-facebooks-machine-learning-algorithms-accurately-predict-suicide/; Mark Zuckerberg, *Building Global Community*, FACEBOOK (Feb. 16, 2017), https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634

Facebook announced its application of AI to identify suicidal intent in user-generated content.[73] According to a company spokesperson, machine learning algorithms scan users' posts, and comments made in response to those posts, for cues that reflect elevated suicide risk.[74] In a Facebook promotional video released on November 26, 2017,the Chautauqua County Sheriff's Department in Upstate New York praises Facebook for alerting it to a potential suicide, which enabled officers to intervene.[75] The following day, Facebook announced its AI-based suicide prediction program had initiated over 100 such "wellness checks," which are often referred to as welfare checks by the law enforcement community. In that announcement, Facebook said it would expand its suicide prediction program globally in "most of the countries in which it operates, with the exception of those in the European Union (EU)."[76] In contrast to the soft-touch interventions implemented by the VA through REACH VET, sending police to people's homes can be thought of as a firm-hand intervention because it is relatively invasive, done without people's consent, and Facebook users cannot refuse to speak with emergency responders.

On April 2, 2018, Zuckerberg revealed that Facebook's AI scans the contents of users' private messages, which suggests that both public and private user-generated content may be scanned for signs of suicidal intent.[77] On September 10, 2018, Facebook provided additional details about its suicide prediction algorithms: Using a AI tool called random forests, Facebook analyzes user-generated content and assign a risk-rating to words, word pairs, and phrases in each post. Hypothetical examples provided by the company include "sadness," "much sadness," and "so much sadness." This method is like the approaches used by Walsh and Poulin. However, in Facebook's case, the words and phrases are derived from social media content instead of medical records. Like Walsh's model, Facebook's system produces a score that ranges from zero to one, where one represents the highest risk of imminent suicide.[78]

Less is known about Facebook's use of AI to predict suicide on its photo sharing site Instagram. However, Facebook has developed sophisticated AI that can identify the content of images.[79] Called "computer vision," this technology could be trained to identify objects that are associated with suicide such as firearms, pill bottles, and the ledges of buildings or the railings on bridges

---

[73] Kwon *supra* note 39; Vanessa Callison-Burch, *Building a Safer Community with New Suicide Prevention Tools,* Facebook (Mar. 1, 2017), https://newsroom.fb.com/news/2017/03/building-a-safer-community-with-new-suicide-prevention-tools/.
[74] Kwon *supra* note 39.
[75] Facebook Safety, https://www.facebook.com/fbsafety/videos/1497015877002912/
[76] Hayley Tsukayama, *Facebook is Using AI to Try to Prevent Suicide*, WASH. POST (Nov. 27, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/11/27/facebook-is-using-ai-to-try-to-prevent-suicide/?noredirect=on&utm_term=.55095e182542.
[77] Sarah Frier, *Facebook Scans the Photos and Links You Send on Messenger,* BLOOMBERG (Apr. 4, 2018), https://www.bloomberg.com/news/articles/2018-04-04/facebook-scans-what-you-send-to-other-people-on-messenger-app.
[78] Benjamin Goggin, *Inside Facebook's suicide algorithm: Here's how the company uses artificial intelligence to predict your mental state from your posts,* BUS. INSIDER (Jan 6. 2019), https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12.
[79] *Understanding the visual world around us*, Facebook Research, https://research.fb.com/category/computer-vision/.

and balconies.[80] Independent researchers have previously demonstrated that AI can analyze Instagram posts to infer users' moods and whether they are depressed.[81]

Unlike medical suicide prediction, which is mostly experimental, transparent, subject to health laws, and approved by IRBs, Facebook's suicide prediction programs are not subject to these rules and regulations, and its methods and outcomes are unpublished. This lack of transparency and accountability raises safety concerns that are discussed in Part II. Instead of consulting an IRB, Facebook sometimes utilizes an internal "ethics board."[82] However, unlike customary IRB approval, which is mandated by the Common Rule, review of Facebook's projects by its ethics board occurs at the company's discretion.[83]

Facebook's lack of transparency is concerning because the company has a history of monitoring people's emotional states and experimenting on users without their knowledge or consent.[84] Since the company's wellness checks were made public in late 2017, Facebook has expanded its suicide prediction program internationally and conducted at least 3,500 wellness checks in the US and abroad.[85] However, many questions remain unanswered. For example, on what data were its algorithms trained? Facebook provides only vague answers. According to an article written by its software engineers: "To start, we worked with experts to identify specific keywords or phrases known to be associated with suicide."[86] However, Facebook quickly learned this approach resulted in too many false positives, picking up benign phrases such as "Ugh, I have so much homework I just wanna kill myself," which is meant to express frustration rather than suicidal intent.[87]

Facebook then implemented an AI-based approach using machine learning. According to its engineers: "We were able to use posts previously reported to Facebook by friends and family, along with the decisions made by our trained reviewers (based on our Community Standards), as our training data set."[88] This quote reveals a serious limitation of Facebook's AI training method. Because the company lacks access to medical records, it cannot train its AI using data from actual suicides. Instead, it appears to use the reports of concerned Facebook users and the subsequent actions of its content moderators as a proxy for suicide risk. Facebook's approach has severe limitations because instead of accurately predicting suicidal thoughts and behaviors,

---

[80] *See Id.*
[81] Andrew G. Reece and Christopher M. Danforth, *Instagram photos reveal predictive markers of depression*, 6 EPJ DATA SCI. (2017).
[82] Molly Jackman and Lauri Kanerva, *Evolving the IRB: Building Robust Review for Industry Research,* 72 WASH. & LEE L. REV. ONLINE 422 (2017).
[83] *Id.; See* also 45 C.F.R. §46.109 (2018).
[84] *See* Robinson Meyer, *Everything We Know About Facebook's Secret Mood Manipulation Experiment,* ATLANTIC (Jun. 28, 2014), https://www.theatlantic.com/technology/archive/2014/06/everything-we-know-about-facebooks-secret-mood-manipulation-experiment/373648/; *See* also Sam Levin, *Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless,'* GUARDIAN (May 1, 2017), https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens.
[85] Kaste *supra* note 7.
[86] Dan Muriello et al., *Under the hood: Suicide prevention tools powered by AI,* FACEBOOK CODE (Feb. 21, 2018), https://code.fb.com/ml-applications/under-the-hood-suicide-prevention-tools-powered-by-ai/.
[87] *Id.*
[88] *Id.*

Facebook’s AI may merely be predicting what its users and content moderators perceive to be suicide risk.

In an e-mail interview, I asked Facebook’s Emily Cain whether the company retains information about the outcomes of wellness checks. She said: “Most of the time we do not know the outcome of those wellness checks because first responders usually keep that information confidential. On occasion, first responders will respond to Facebook's escalation to share the outcome of the intervention.” Thus, Facebook may receive some suicide data from emergency responders following wellness checks. However, it is unknown whether the company feeds this data back into the system to improve its suicide predictions.

A lack of real-world suicide data would significantly reduce the accuracy of Facebook’s predictions. I asked whether the company conducts experiments to test their accuracy. According to Cain, “We audit and perform quality checks to ensure that we’re moderating content for suicide and self injury appropriately (taking down violating content, checkpointing people who appear to be in crisis, escalating imminent issues) the same way we do across all content on Facebook.”[89] However, she declined to provide further information regarding these processes.

When asked what training and certification Facebook’s content moderators have and what criteria they use to decide when police should be contacted, Cain responded:

> Our Community Operations team includes thousands of people around the world who review reports about content on Facebook. The team includes a dedicated group of specialists who have specific training in suicide and self harm . . . Where we have signals of potential imminent risk or harm, a specialized team conducts an additional review to determine if we should help refer the individual for a wellness check. Those teams are trained to engage directly with first responders to assist them in locating the person to conduct a wellness check. This team has experience in safety, law enforcement response, or crisis response with backgrounds in domestic and federal U.S law enforcement, rape and suicide hotlines, Center for Missing or Sexually Exploited Children, Social Services, international law enforcement as well as domestic and international crisis and intervention centers.

It may seem reassuring that Facebook’s community operations team includes people with experience working in crisis intervention. However, without more information about their credentials and how they make decisions, it is difficult to evaluate the safety, fairness, and efficacy of the program.

b. Crisis Text Line

Launched in August of 2013, Crisis Text Line is a text-based crisis support service marketed to children and teens. It aims to reach them through texting, the communication medium they use most. Based on 65 million text messages, the company reports it has identified over ten thousand words, word pairs, and triplets that “were a better indicator of a suicide risk attempt than the

[89] E-mail interview between Mason Marks and Facebook’s Emily Cain (Nov. 28, 2018).

word suicide itself . . . we used . . . what's called a deep neural net to power this algorithm and identify these words and phrases." According to Crisis Text Line, the word "military" reflects twice the risk of a suicide attempt than the word suicide alone, the crying face emoticon carries a suicide risk eleven times that of the word suicide, and words for over-the-counter medicines such as ibuprofen and Excedrin carry a risk fifteen times higher than the word suicide. Using its algorithm, Crisis Text Line claims it can identify 86% of high-risk texters based on a single text message.[90] However, this claim cannot be verified because Crisis Text Line has not made its data public or explained how it defines "high-risk texters." Crisis Text Line's AI training process appears to suffer from the same drawbacks as Facebook's. Because Crisis Text Line lacks access to real suicide data, like Facebook, it relies on its past internal actions to train its AI.[91] In this case, the company's training data appears to consist of past decisions made by its crisis counselors in response to concerning texts.[92] In other words, like Facebook, Crisis Text Line appears to use the past decisions of its employees as a proxy for suicide risk.

Crisis Text Line is more forthcoming than Facebook about the demographics of people affected by its suicide prediction algorithms. Like Facebook, it contacts emergency responders to perform wellness checks, which it calls "active rescues," on users deemed high risk for self-harm. According to the company's Chief Data Scientist: "An active rescue is when we send out emergency services to intervene in an active suicide attempt. We are doing over 20 of these a day right now."[93] The company claims to have completed over 11,500 active rescues.[94] Crisis Text Line serves a population that contains a higher percentage of vulnerable groups than the general population. For example, 45% of its users identify as LGBTQ, 20% as Hispanic, and 5% as Native American.

According to Founder Nancy Lublin: "It turns out our texters skew young, poor, and rural. 75% of our users are under age 25 including . . . 10% of our users under the age of 13 . . . if you take the nation's lowest 10% by socioeconomic status area codes, that 10% is using 19% of our volume. So we double over-index the poorest people in America . . . rural area codes, rural locations where they don't have access to mental health and behavioral health services including 5% of our texters indicate that they are native American or native Alaskan, which is interesting because only 1.5% of America identifies that way."[95] These demographics are important because they show that Crisis Text Line's user base is composed largely of people from marginalized groups. Without oversight and persuasive evidence that suicide predictions safe, fair, and effective, members of these vulnerable groups may be disproportionately impacted by the risks associated with social suicide prediction, which are discussed further in Part II. However, little is known about how Crisis Text Line's algorithms were designed, what data they were trained on, whether they are safe and effective, what outcomes resulted from its more than 11,500 wellness checks, and who has access to its suicide prediction data.

---

[90] Wharton School, *Wharton People Analytics Conference 2018: Social Impact Perspective: Bob Filbin*, YOUTUBE (May 9, 2018), https://www.youtube.com/watch?v=e3WWCDFQqmA.
[91] *Id.*
[92] *Id.*
[93] *Id.*
[94] *Id.*
[95] O'Reilly, *Crisis Text Line Data Usage and Insights – Nancy Lublin & Bob Filbin (Crisis Text Line)*, YOUTUBE (Mar. 27, 2018) https://www.youtube.com/watch?v=DBY_j77_Ehc.

In 2018, Crisis Text Line revealed it shares counseling data with a for-profit spinoff called Loris.AI.[96] Though it is not illegal for non-profit corporations to own for-profit subsidiaries, the commercialization of mental health and suicide-related data from vulnerable populations, including children under the age of 13, deserves scrutiny.

Crisis Text Line has formed partnerships with Facebook, YouTube, and the developers of text messaging apps targeted at teens such as After School and Kik. [97] Users of these apps can access Crisis Text Line counselors through each platform's interface. In May of 2017, Crisis Text Line formed a partnership with the California Community College system, composed of 114 two-year colleges, to provide crisis support to its students.[98] Other partners in higher education include Iowa State, Penn State, and the University of San Francisco.[99]

When college students use Crisis Text Line, the topics of conversation are coded and recorded along with other data such as each user's area code.[100] This information is shared with California community college administrators.[101] Though Crisis Text Line claims that its texters remain anonymous, the practice of reporting texter data to university officials raises privacy concerns. It is well established that "de-identified data" can often be re-identified with only a few additional pieces of information.[102] Having students' area codes and topics of conversation may be enough to re-identify them when combined with other data from their educational records.

According to Crisis Text Line's west coast director, California community college students are four times more likely to discuss homelessness, and three times more likely to discuss financial problems, with the text line than its other users nationwide.[103] The demographics of Crisis Text Line's users will come into play in Part II on the risks of social suicide prediction. Specifically, people are at high risk for suicide shortly after being released from psychiatric hospitals.[104] Those who lack access to mental health resources and other support systems may be particularly vulnerable.[105] People in marginalized groups and lower socioeconomic classes may lack access to adequate support following hospitalization, and therefore, may be at greater than average risk following wellness checks that results in involuntary hospitalization.

[96] Sandra Upson, *Can Crisis Line Messaging Help Improve Workplace Culture?,* WIRED (Feb. 6, 2018), https://www.wired.com/story/can-crisis-line-messaging-help-improve-workplace-culture/.
[97] Clinton Nguyen, *This text-message hotline can predict your risk of depression or stress*, BUSINESS INSIDER (Jun. 21, 2016), https://www.businessinsider.com/crisis-text-line-is-gathering-data-about-depression-stress-2016-6.
[98] Adolfo Guzman-Lopez, *New Crisis Text Line Identifies California College Student Homelessness as Big Issue*, SOUTHERN CALIFORNIA PUBLIC RADIO (Sep. 1, 2017), https://www.scpr.org/news/2017/09/01/75231/new-crisis-text-line-identifies-california-college/.
[99] *Id.*
[100] *Id.*
[101] *Id.*
[102] *Id.*
[103] *Id.*
[104] Daniel Thomas Chung et al., *Suicide Rates After Discharge from Psychiatric Facilities*, 74 JAMA PSYCHIATRY 694 (2017); Mark Olfson et al., *Short-term Suicide Risk After Psychiatric Hospital Discharge,* 73 JAMA PSYCHIATRY 1119 (2016).
[105] *See Id.*

c. Operation Zero

The approaches described so far have focused on the analysis of text and video to predict suicide risk. However, suicide prediction is not limited to these approaches. A company called Operation Zero is experimenting with GPS-derived location data to predict depression and suicidal thoughts and behaviors.[106] Companies routinely use location data to track consumers and learn their habits and preferences.[107] A recent New York Times article revealed how GPS tracking can be used to monitor consumers and identify the stores they visit (and even which parking spaces they use).[108] The information revealed is often sensitive: one example involves a consumer who visited a Planned Parenthood office for two hours.[109]

Operation Zero plans to use location data, derived using technology developed by Foursquare, to track the movements of veterans who download the company's app and identify patterns that reflect mental illness and suicidal thoughts.[110] In one example, a veteran who is usually physically active stops going to the gym (as reflected by GPS tracking), which Operation Zero claims might reflect the onset of depression.[111] Location data might also track individuals who travel to locations that are common sites for suicide such as the Golden Gate Bridge and China's Nanjing Yangtze River Bridge.[112] In the future, location data might be cross-references with other information, for example from social media, to calculate suicide risk scores, and police might be notified when people categorized as high risk approach or linger at locations that are common sites for suicide. Authorities might also be notified when people labeled high risk visit gun stores or purchase ammunition and other items associated with suicide.[113]

d. Google, YouTube, and Google Assistant

The extent to which Google conducts social suicide prediction in unknown. However, when users enter suicide-related terms into Google's search engine, its AI identifies the nature of their inputs and provides them with resources such as phone numbers for suicide hotlines (Yahoo,

---

[106] Jesse L, *Announcing the winner of our first 'Foursquare for Good' program,* MEDIUM (Nov. 27, 2018), https://medium.com/foursquare-direct/announcing-the-winner-of-our-first-foursquare-for-good-program-c512f62e966e.
[107] Jennifer Valentino-DeVries et al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, NY TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.
[108] *Id.*
[109] *Id.*
[110] Jesse L *supra* note.
[111] *Id.*
[112] *See* Neil Tweedie, *Golden Gate Bridge is the world's most popular site for suicide: 'Just why do they make it so easy?',* TELEGRAPH (May 26, 2012), https://www.telegraph.co.uk/news/features/9289970/Golden-Gate-Bridge-is-the-worlds-most-popular-site-for-suicide-Just-why-do-they-make-it-so-easy.html; Jennifer Chaussee, *Mesh Nets to Catch Suicide Jumpers May Be Placed Under the Golden Gate Bridge,* BUS. INSIDER (Jun. 27, 2014), https://www.businessinsider.com/r-nets-to-catch-suicide-jumpers-may-be-placed-beneath-iconic-golden-gate-bridge-2014-26 (reporting that the Nanjing Yangtze River bridge is the word's most popular site of suicide attempts, and the Golden Gate Bridge is the second most popular).
[113] *See* Andrew Ross Sorkin, *How banks unwittingly finance mass shootings,* NY TIMES (Dec. 24, 2018), https://www.nytimes.com/interactive/2018/12/24/business/dealbook/mass-shootings-credit-cards.html.

Bing, and Facebook searches yield similar results).[114] Outside of Google, independent researchers have attempted to predict suicide trends based on regional Google searches with mixed results.[115]

In addition to its search engine, Google's other Internet platforms could generate suicide predictions. For instance, AI might analyze the text of Gmail messages to infer suicidal thoughts and predict suicide attempts.[116] Google likely already uses AI to identify suicide-related behavior on its streaming platform YouTube.[117]

Google is moving aggressively into the hardware and IoT space. In 2018, the company obtained a patent for a "smart home" capable of making health-related inferences about its inhabitants.[118] One embodiment of the invention can infer substance use, such as alcohol or tobacco consumption, and infer medical conditions such as Alzheimer's disease, by analyzing chemical traces, audio, video & household occupants' movement patterns.[119] This embodiment of Google's invention does not necessarily reflect an existing or planned Google product because companies often patent inventions that do not yet exist. However, the patent demonstrates how the company's technology could potentially be used to infer mental health conditions and suicide risk. Google could conceivably combine behavioral data collected by smart homes and other Internet of things devices with insights gleaned from its research with medical records to make powerful health-related inferences.

e. Amazon Alexa

Amazon leads the market for personal digital assistants.[120] IoT devices featuring Amazon Alexa are the bestselling models with over 100 million units sold, and the technology is being incorporated into products as diverse as microwaves, cars, and security cameras.[121] These products will collect an unprecedented volume of speech and other behavioral data from

---

[114] Lucas Chae, *How search engines are failing suicidal users,* FAST COMPANY (Sep. 6, 2018), https://www.fastcompany.com/90230313/how-search-engines-are-failing-suicidal-users.
[115] *See* Michael J. McCarthy, *Internet Monitoring of Suicide Risk in the Population,* 122 J. AFFECTIVE DISORDERS 277 (2010); *see* also Vitaliy Bezsheiko, *Google Trends as a method for prediction of suicide rates,* 2 Psychosomatic Medicine and General Practice (2017); *see* also Ulrich S. Tran et al., *Low validity of Google Trends for behavioral forecasting of national suicide rates,* 16 PLOS ONE 21 (2017).
[116] *See* Douglas MacMillan, *Tech's 'Dirty Secret': The App Developers Sifting Through Your Gmail,* Wall St. J. (Jul. 2, 2018), https://www.wsj.com/articles/techs-dirty-secret-the-app-developers-sifting-through-your-gmail-1530544442.
[117] *See* Louise Matsakis, *A Window Into How YouTube Trains AI to Moderate Videos,* WIRED (Mar. 22, 2018), https://www.wired.com/story/youtube-mechanical-turk-content-moderation-ai/.
[118] U.S. Patent No. 10,114,351 B2 "Smart-home automation system that suggests or automatically implements selected household polities based on sensed observations."
[119] *Id.*
[120] Karen Hao, *Amazon Echo's dominance in the smart-speaker market is a lesson on the virtue of being first,* QUARTZ (Jan. 8, 2018), https://qz.com/1157619/amazon-echos-dominance-in-the-smart-speaker-market-is-a-lesson-on-the-virtue-of-being-first/.
[121] Lucas Matney, *More than 100 million Alexa devices have been sold,* TECHCRUNCH (Jan. 4, 2019), https://techcrunch.com/2019/01/04/more-than-100-million-alexa-devices-have-been-sold/; Bloomberg, *Amazon Just Unveiled a Bunch of New Stuff With Alexa – Including a $60 Microwave,* TIME (Sep. 20, 2018), http://time.com/5402499/amazon-alexa-microwave/.

consumers. According to Alexa chief scientist Rohit Prasad: "Amazon's Alexa team is beginning to analyze the sound of users' voices to recognize their mood or emotional state."[122] Amazon could combine data collected by Alexa-enabled devices with its growing database of medical records, for example from Amazon Comprehend Medical, to train its AI to make social suicide predictions.

f. Twitter

The extent to which Twitter uses AI to predict suicidal thoughts and behaviors is unclear. However, several independent teams have conducted suicide research using publicly available Twitter posts. In 2014, a British suicide prevention group called the Samaritans introduced an app that notified Twitter users when people they followed posted concerning Tweets.[123] The app tracked the activity of nearly 1.9 million users.[124] Critics are concerned the system could mislabel millions of people and promote discrimination against those with mental illness.[125] In Canada, Zachary Kaminski at the Royal Ottawa Hospital is developing software to predict suicide by analyzing tweets. He reports it can predict suicide with 89 percent accuracy but cautions there is a lot of room for improvement.[126]

Researchers in China have developed an algorithm for predicting suicide in users of Weibo, a popular Chinese messaging app, and they are producing an English version for Twitter users.[127] When developing the algorithm, the team approached Weibo and proposed a partnership to integrate suicide predictions into the platform.[128] When the company declined, the team conducted suicide predictions independently using publicly available Weibo data. When the researchers identified high risk users, they contacted the users directly through Weibo's messaging feature and referred them to mental health resources.[129] The team is now working with the University of Maryland and Brigham Young University to adapt its prediction model for the Twitter platform.[130]

g. Public Health Agency of Canada

In early 2018, the Public Health Agency of Canada announced it had hired an Ottawa-based company called Advanced Symbolics to develop social suicide prediction software.[131] Advanced Symbolics specializes in political predictions and is best known for accurately predicting Brexit and the election of Donald Trump. Unlike other social suicide prediction

---

[122] Khari Johnson, *Amazon's Alexa wants to learn more about your feelings,* VENTURE BEAT (Dec. 22, 2017), https://venturebeat.com/2017/12/22/amazons-alexa-wants-to-learn-more-about-your-feelings/.
[123] Natasha Singer, *Risks in Using Social Media to Spot Signs of Mental Distress,* NY TIMES (Dec. 26, 2014), https://www.nytimes.com/2014/12/27/technology/risks-in-using-social-posts-to-spot-signs-of-distress.html.
[124] *Id.*
[125] *Id.*
[126] *Id.* (Kaminsky on YouTube: https://www.youtube.com/watch?v=0uFPpG-2QC8)
[127] Mandy Zuo, *How China's AI technology can help Twitter's suicidal users,* SOUTH CHINA MORNING POST (Feb. 3, 2018), https://www.scmp.com/news/china/society/article/2131853/how-chinese-ai-technology-may-help-find-suicidal-posts-twitter.
[128] *Id.*
[129] *Id.*
[130] *Id.*
[131] https://www.macleans.ca/society/how-ai-is-helping-to-predict-and-prevent-suicides/

programs, which predict suicide in individuals, the Advanced Symbolics will identify suicide trends at the population level.[132]

h. U.S. Department of Veterans Affairs Durkheim Project

In addition to its REACH VET medical suicide prediction program, the VA has engaged in social suicide prediction through an opt-in program for veterans called the Durkheim Project, which ran between 2011 and 2015. Unlike Facebook users, who are unable to opt-out of the company's suicide prediction program, veterans at the VA had to opt-in to being tracked through the Durkheim Project. The project used Facebook to recruit research subjects.[133] When Facebook identified users who were veterans or active military personnel, they were shown a pop-up inviting them to participate in the study.[134] After opting-in, the Durkheim Project's software monitored veterans' social media accounts including Facebook, Twitter, LinkedIn, and the now defunct Google Plus.[135]

Clinicians could log-in to a dashboard to view each veteran's suicide risk, which was displayed using a color-coded scale in which green represented low risk, yellow indicated nominal risk, and red indicated high risk.[136] An expanded view displayed a numerical risk score and a probability (shown as a percentage) indicating the degree of confidence in the score.[137] Trends in suicide prediction for each patient could be viewed, and one section of the dashboard showed each veteran's suicide risk relative to the clinician's other patients.[138] Clinicians could also view the original social media content that formed the basis for each risk calculation.[139]

The Durkheim Project was conducted in three phases. In the first phase, which was described above in the section on medical suicide prediction, researchers used a sample of unstructured clinical notes derived from VA medical records to develop a linguistics-based model for predicting suicide risk.[140] In the second phase, researchers used this model to analyze social media content and calculate suicide risk in veterans and military personnel who opted-in to the study.[141] Here it is important to point out the differences between the approach taken by the Durkheim Project and the approaches of Facebook and Crisis Text Line. Whereas Facebook and Crisis Text Line lack access to medical records, and they must use proxies for suicide to train their AI suicide prediction software, researchers at the Durkheim Project had access to VA medical records, and they used data from actual suicides to train their suicide prediction algorithms. The researchers acknowledge one important shortcoming to their approach: the text on which their algorithms were trained (unstructured clinical notes) was written by physicians, whereas the social media content on their suicide predictions were based was written by

---

[132] Sydney Kennedy and Trehani M. Fonseka, *How AI is helping to predict and prevent suicides*, MACLEANS (Mar. 29, 2018), https://www.macleans.ca/society/how-ai-is-helping-to-predict-and-prevent-suicides/.
[133] Poulin *supra* note 23.
[134] *Id.*
[135] Durkheim Project Demo, http://durkheimproject.org/demo/
[136] *Id.*
[137] *Id.*
[138] *Id.*
[139] *Id.*
[140] Poulin *supra* note 23.
[141] *Id.*

veterans. There will inevitably be differences in the words and phrases used by healthcare providers to describe their patients, and the words used by veterans on social media. However, the researchers hypothesize that there is overlap between the words and phrases used in these contexts. For example, words like “anxious” or “anxiety” might be used both by veterans on social media to explain how they are feeling and by physicians caring for those veterans to describe their mental state. Physicians might also quote veterans in their clinical notes. Despite any potential shortcomings of this hybrid approach, the methods used by the Durkheim Project are grounded in the scientific method and their predictions are based on data from real suicides.

There are other important differences between the Durkheim Project and the suicide prediction programs at Facebook and Crisis Text Line. The Durkheim Project gave healthcare providers a clinical dashboard allowing them to view the social media content that resulted in each patient’s suicide risk score. This feature allowed providers to use their clinical judgment to decide whether additional follow-up was necessary. In contrast, we do not know what information Facebook and Crisis Text Line provide to emergency responders. Do they provide police with transcripts of the relevant social media content allowing police to view it in context, or do they simply relay that the user is at high risk for suicide? Without greater transparency, it is impossible to know. In addition, unlike social suicide predictions made by Facebook and other social media platforms, those made through the Durkheim Project were subject to health laws such as HIPAA because predictions were made by covered entities including the VA and its partners the Geisel School of Medicine at Dartmouth College and the Dartmouth-Hitchcock Medical Center.[142] These differences are discussed further in the following section.

II. AI-based Suicide Prediction Poses Risks to Patients and Consumers

AI holds promise for improving suicide predictions. However, it exposes people to a variety of dangers, which can be divided into safety, privacy, and autonomy risks: Safety risks include false negatives that may leave suicidal individuals without assistance, false positives that can cause biased treatment by physicians, unexpected and unwarranted visits from police that may escalate to violent confrontations, and involuntary medical treatment; Privacy risks include the leak of sensitive information through security breaches, and the transfer or sale of personal data to third parties such as data brokers and advertisers, which can lead to stigmatization, exploitation, and discrimination; and, Autonomy risks include censorship, unnecessary confinement or civil commitment, and in countries where suicide attempts are illegal, criminal penalties including fines and incarceration. The following sections describe these risks in greater detail. There is significant overlap between the risks of medical and social suicide prediction. For this reason, the risks of both categories are discussed together. However, because medical suicide prediction is governed by health laws and regulations, people subjected to it are provided greater protection than those who are subjected to social suicide prediction.

1. Safety Risks

The safety risks of AI-based suicide predictions stem from their inaccuracy and the limited effectiveness of interventions that are triggered by predictions. Despite purported improvements

[142] Durkheim Project Demo, http://durkheimproject.org/demo/

over traditional prediction methods, AI-based predictions produce many false positives and false negatives.[143] Both types of misclassification can affect people's safety. The risks associated with false negatives are easiest to understand. If suicide predictions are less than 100 percent accurate, they will inevitably fail to identify some suicidal people. Those individuals might not receive needed assistance and may harm or kill themselves.

By comparison, the safety risks of false positives are more complex. They stem from stigmatization and the treatment interventions that result from being labeled high-risk for suicide. People placed in this category may be treated differently by physicians in ways that endanger their health and safety. Dr. Greg Simon likens false positives from suicide prediction algorithms to false positives from vehicle blind spot warning systems. If blind spot warning system issues false positives, the driver can act as though they are true positives, postpone switching lanes, and little harm is done. In the worst-case scenario, he might miss his offramp and have to double back. This analogy may hold true in limited cases. For instance, if the result of a false positive is a non-invasive, soft-touch intervention, the harm to a patient or consumer may be minimal. However, for the most part, Simon's analogy is a poor fit for suicide predictions. If suicide screening tools produce false positives, there may be long-lasting and potentially fatal adverse effects.

According to an article in the British Journal of Psychiatry: "The most obvious harm is that patients labelled 'high risk' may receive needlessly more restrictive treatments."[144] For example, patients might be taken off certain medications due to the perceived suicide risk even if the medications are helpful. One current example involves opioids. Despite the ongoing US opioid crisis, opioids remain an appropriate treatment for many patients.[145] However, in the context of the crisis, physicians are increasingly reluctant to prescribe opioids, and if an algorithm labels a patient high risk for suicide, doctors might respond by withholding access to opioids due to the perceived risk of overdose.[146] Patients undergoing surgery may receive inadequate post-operative pain control, and patients prescribed opioids for chronic pain may be abruptly tapered off them. Thus, patients could unnecessarily be forced to endure pain and its complications due to inaccurate suicide predictions. Withdrawing adequate pain control may be inappropriate even if suicide predictions are accurate because many suicides have been blamed on physicians' tapering or withholding opioids resulting intractable pain.[147]

---

[143] Mulder *supra* note 14.
[144] *Id.*
[145] *See* Marilyn Serafini, *The Physicians' Quandary with Opioids: Pain versus Addiction,* NEJM CATALYST (Apr. 26, 2018), https://catalyst.nejm.org/quandary-opioids-chronic-pain-addiction/.
[146] *See* Juliann Garey, *When Doctors Discriminate*, NY TIMES (Aug. 10, 2013), https://www.nytimes.com/2013/08/11/opinion/sunday/when-doctors-discriminate.html (reporting physician bias and refusal to prescribe pain medication following disclosure of bipolar disorder diagnosis).
[147] Thomas Kline, *#OpioidCrisis Pain Related SUICIDES associated with forced tapers,* MEDIUM (May 11, 2018), https://medium.com/@ThomasKlineMD/opioidcrisis-pain-related-suicides-associated-with-forced-tapers-c68c79ecf84d; Elizabeth Llorente, *As doctors taper or end opioid prescriptions, many patients driven to despair, suicide,* FOX NEWS (Dec. 10, 2018), https://www.foxnews.com/health/as-opioids-become-taboo-doctors-taper-down-or-abandon-pain-patients-driving-many-to-suicide.

Due to false positives, patients might be hospitalized against their will, and a diagnosis of suicidal thoughts would become part of their permanent medical records. Healthcare providers may find it difficult to ignore the results of AI-based suicide predictions even when they disagree with the predictions and suspect they might be false positives. Similar concerns have been raised in the context of the justice system, where judges use opaque, proprietary algorithms in sentencing and parole hearings to predict who is likely to recidivate.[148] Though sentencing decisions are ultimately the responsibility of judges, they may be influenced by algorithmic assessments.[149] In the healthcare setting, doctors may be incentivized to follow AI-based suicide predictions because overriding a prediction could expose them to medical malpractice liability if they don't hospitalize patients who subsequently attempt or complete suicide.

Involuntary hospitalization and forced medication are not without risks. Though they can prevent suicide in the short term, unnecessary confinement and treatment may paradoxically increase suicide risk because the experience can be traumatic and dehumanizing.[150] In a highly publicized case from 2015, Sandra Bland died by suicide after being confined for three days after a routine traffic stop led to her arrest and imprisonment.[151] Though Bland's arrest, confinement, and suicide following a moving violation is not a perfect analogy to involuntary confinement following a suicide prediction, the sequence of events illustrates how involuntary confinement can be traumatic and increase one's suicide risk.

People are often at increased risk for suicide shortly after being admitted to hospitals and shortly after being released.[152] Moreover, it is well documented that numerous psychiatric medications are associated with transient increases in suicide risk. These risks may be exacerbated when people lack access to mental health resources and social support outside the hospital. The Crisis Text Line user demographics described above demonstrate that crisis counseling services may be disproportionally utilized by vulnerable populations including the homeless, certain racial and sexual minorities, and people from the lowest-income zip codes. Individuals in these groups may lack adequate social, medical, and psychological support outside the hospital, and they may be particularly vulnerable to the risks associated with involuntary hospitalization and forced medication.

There is also a risk that doctors will treat patients categorized as high-risk differently than other patients. Physicians are sometimes biased against patients with mental illnesses, substance use disorders, and histories of suicidal thoughts.[153] As a result, a false positive placed into a patient's

---

[148] *See* Jason Tashea, *Courts are Using AI to Sentence Criminals. That Must Stop Now,* WIRED (Apr. 17, 2017), https://www.wired.com/2017/04/courts-using-ai-sentence-criminals-must-stop-now/.
[149] *Id.*
[150] *See* B. Olofsson and L. Jacobsson, *A plea for respect: involuntary hospitalized psychiatric patients' narratives about being subjected to coercion,* 8 PSYCHIATRIC MENTAL HEALTH NURSING 357 (2001); *see* also Gail C. Eisenberg, *Involuntary Commitment and the Treatment Process: A Clinical Perspective,* 44 BULLETIN AMER. ACAD. PSYCHIATRY L. (1980).
[151] https://www.washingtonpost.com/news/morning-mix/wp/2015/07/22/documents-sandra-bland-previously-attempted-suicide-felt-very-depressed-on-day-of-arrest/?utm_term=.845c88f9fd34
[152] *See* Ping Qin and Merete Nordentoft, *Suicide Risk in Relation to Psychiatric Hospitalization,* 62 ARCH. GEN. PSYHICATRY 427 (2005).
[153] Stephanie Knaak, *Mental illness-related stigma in healthcare,* 30 HEALTHCARE MANAGEMENT FORUM 111 (2017).

record may affect how physicians treat the patient in the future. Patients with mental illnesses often report feeling dehumanized and dismissed by healthcare providers.[154]

Healthcare providers, social media platforms, and police may over rely on suicide predictions. According to one metanalysis on suicide risk assessment: "[A]n over-reliance on the identification of risk factors in clinical practice, is, in our view, potentially dangerous and may provide a false reassurance for clinicians and managers."[155] The authors emphasize that clinicians should draw a distinction between risk assessment and prediction: "The idea of risk assessment as risk prediction is a fallacy and should be recognized as such. We are simply unable to say with any certainty who will and will not go on to have poor outcomes. People who self-harm often have complex and difficult life circumstances, and clearly need to be assessed- but we need to move away from assessment models that prioritise risks at the expense of needs."[156] Thus, the authors appear to advocate for a soft-touch approach to suicide prevention in which risk assessments lead to more thorough, individualized evaluations instead of firm-hand suicide interventions.

Firm-hand interventions such as sending police to people's homes could have unexpected consequences such as exacerbation of symptoms and involuntary hospitalization. Police response may further escalate already tense situations. There are numerous reports of people being shot by police after they arrive to investigate erratic behavior or a threat of suicide. In some cases, it is believed suicidal individuals provoke police with the goal of being shot, which is termed "suicide by cop." In other cases, the reasons for police shootings are less clear.

On June 14, 2014, Jason Harrison's mother called Dallas police requesting their help transporting him to a hospital for psychiatric care.[157] Harrison was 38 years old and had been diagnosed with schizophrenia and bipolar disorder.[158] When police arrived, Harrison stood in the doorway holding a small screwdriver.[159] Despite carrying less-than-lethal weapons such as Tasers and pepper spray, two officers drew their firearms and shot and killed Harrison.[160]

On March 9, 2015, an Atlanta-area police officer shot and killed 27 year old Air Force veteran Anthony Hill.[161] According to Hill's family, he was experiencing a non-violent episode resulting

---

[154] Stephanie Knaack, Ed Mantler, and Andrew Szeto, *Mental Illness-related Stigma in Healthcare- Barriers to Access and Care and Evidence-based Solutions,* 30 HEALTHCARE MANAGEMENT FORUM 111 (2017).
[155]
[156] MKY Chan et al., *Predicting suicide following self-harm: systematic review of risk factors and risk scales*, 209 BRITISH J. PSYCHIATRY 277, 279 (2016). [change to *supra* note]
[157] Curtis Skinner, *Family of Jason Harrison, Mentally Ill Man Killed by Dallas Police, Release Graphic Video*, Huffington Post (Mar. 17, 2015), https://www.huffingtonpost.com/2015/03/17/jason-harrison-shooting-v_n_6887242.html.
[158] *Id.*
[159] *Id.*
[160] *Id.* (showing officer with Taser and pepper spray holstered on his utility belt)
[161] Associated Press, *Atlanta-area police officer charged with felony murder for shooting of Anthony Hill*, GUARDIAN (Jan. 22, 2016), https://www.theguardian.com/us-news/2016/jan/22/anthony-hill-shooting-atlanta-georgia-police-felony-murder-charge-robert-olsen.

from trauma endured while on active duty in Afghanistan.[162] Hill had previously been treated for bipolar disorder.[163] On the night of his death, police responded to reports that he had jumped from a second story balcony and was behaving erratically on the grounds of an apartment complex.[164] When police arrived, Hill approached an officer while naked and unarmed.[165] Though the officer carried a Taser, he drew his firearm before shooting and killing Hill.[166]

These examples show that police intervention involving people with mental illness can quickly spiral out of control. Moreover, both Harrison and Hall were black men, and research suggests that black people are more likely to be shot by police than their white counterparts.[167] The difference in police violence is exaggerated when the people shot are unarmed.[168]

Three examples from 2018 illustrate the dangers of relying on third-party reports from social media to initiate wellness checks. In each case, police may have responded with aggression out of proportion to the risk posed to them. On January 20, 2018, high school student John Albers was shot and killed by police responding to a 911 call claiming he threatened to kill himself during a video chat session on Apple's Facetime.[169] The dispatcher informed police that Albers was alone and in the basement of his family's home.[170] According to police, as they approached the home, a garage door opened, and a vehicle emerged and moved towards one officer.[171] The officer fired 13 shots into the family minivan killing Albers.[172] His mother filed a lawsuit claiming the police "acted recklessly and deliberately" by killing Albers while he was "simply backing his mom's minivan out of the family garage."[173]

On May 27, 2018, former Army intelligence analyst Chelsea Manning posted two concerning tweets suggesting she might attempt suicide.[174] A wellness check was initiated when people saw the tweets and contacted police.[175] Surveillance cameras in Manning's apartment building

---

[162] Yanan Wang, *Georgia police officer indicted for murder in shooting of unarmed, naked black veteran*, WASH. POST (Jan. 22, 2016), https://www.washingtonpost.com/news/morning-mix/wp/2016/01/22/georgia-police-officer-indicted-for-murder-in-shooting-of-unarmed-naked-black-veteran/?utm_term=.a05bbf170323#comments.
[163] *Id.*
[164] *Id.*
[165] *Id.*
[166] Ashley Southall, *Naked Black Man Fatally Shot by White Police Officer in Georgia*, NY TIMES (Mar. 9, 2015), https://www.nytimes.com/2015/03/10/us/naked-black-man-fatally-shot-by-white-police-officer-in-georgia.html?_r=0.
[167] German Lopez, *There are huge racial disparities in how US police use force*, VOX (Nov. 14, 2018), https://www.vox.com/identities/2016/8/13/17938186/police-shootings-killings-racism-racial-disparities (reporting an analysis of publicly available FBI data on police shootings).
[168] *Id.*
[169] Joe Robertson and Tony Rizzo, *FaceTime suicide threat led police to OP student's home before officer shot him,* KAN. CITY STAR (Jan. 22, 2018), https://www.kansascity.com/news/local/article196001754.html.
[170] *Id.*
[171] *Id.*
[172] Joe Robertson et al., *Lawsuit: Teen killed by Overland Park police was 'simply backing his mom's minivan,'* KAN. CITY STAR (Apr. 17, 2018), https://www.kansascity.com/news/local/crime/article209113834.html.
[173] *Id.*
[174] Micah Lee and Alice Speri, Police Broke Into Chelsea Manning's Home with Guns Drawn- In a "Wellness Check," INTERCEPT (Jun. 5, 2018), https://theintercept.com/2018/06/05/chelsea-manning-video-twitter-police-mental-health/.
[175] *Id.*

recorded the event; the video shows three officers enter Manning's apartment with guns drawn while one officer enters pointing a Taser.[176] The video illustrates how a suicide-related wellness check can escalate to a show force without provocation by a suicidal individual.[177]

The wellness checks describe above were performed in the US. However, in other regions, such as the Middle East and Southeast Asia, police response may be more unpredictable, and wellness checks may result in criminal penalties such as fines and incarceration. Facebook has deployed its suicide prediction system in nearly every region in which it operates except in the European Union. In some countries, attempted suicide is a criminal offense. For instance, in Singapore, where Facebook maintains its Asia-Pacific headquarters, suicide attempts are punishable by imprisonment for up to one year. Attempted suicide is also illegal in nearby Malaysia, Myanmar, and Brunei.[178] In Islamic countries such as Saudi Arabia, Shari'ah law forbids suicide, which is considered a criminal act.[179] In these countries, Facebook-initiated wellness checks might result in criminal prosecution and incarceration.

The above examples illustrate how social suicide prediction is analogous to predictive policing. If Facebook's AI misclassifies a user as suicidal, police could be sent to the person's home, which could escalate the situation and provoke a violent confrontation, involuntary hospitalization, or incarceration. Once police arrive following a report that a person is at high risk for suicide, it may be difficult to convince them to leave without being detained. In one case in Ohio, police detained a woman after Facebook warned law enforcement that she might be suicidal.[180] When police arrived, the woman denied having suicidal thoughts, but the officers informed her she would be transported to a hospital against her will if she refused to comply.[181]

2. Privacy Risks

The privacy risks of suicide prediction stem from how prediction data is stored and where the information flows after predictions are made. The risks include leaking of sensitive information through data breaches, and the transfer or sale of personal data to third parties such as data brokers, lenders, employers, and insurance companies. Sale of suicide-related data to these groups can result in stigmatization, exploitation, and discrimination against people categorized as high risk regardless of whether those categorizations are accurate. For instance, a life insurance company might purchase suicide prediction data on consumers, and then deny them policies or charge them higher rates than individuals with lower suicide risk scores. In 2017, the US Department of Housing and Urban Development (HUD) filed a complaint against Facebook alleging the company violated the Fair Housing Act by allowing advertisers to exclude people with disabilities, and members of some religious faiths and minority groups, from receiving

---

[176] *Id.*
[177] *Id.*
[178] Brian L. Mishara and David N. Weisstub, *The Legal Status of Suicide: A Global Review,* 44 INT'L J. L. PSYCHIATRY 54, 55 (2016).
[179] Mohammed Madadin et al., *Suicide Deaths in Dammam, Kingdom of Saudi Arabia: Retrospective Study,* 3 EGYPTIAN J. FORENSIC SCI. 39, 40 (2013).
[180] Singer *supra* note 111.
[181] *Id.*

housing-related ads.[182] Suicide risk scores could similarly be used to deny people access to housing, employment, and other resources, which might further marginalize this already vulnerable population.

In the healthcare system, HIPAA protects patient privacy, and suicide-related data cannot leave the system without first being deidentified. Healthcare providers are also prohibited from sharing non-anonymized health information with third party advertisers. Thus, medical suicide predictors cannot legally share individualized suicide predictions for marketing purposes. However, because most social suicide predictors are not covered entities under HIPAA, their suicide predictions can be shared with third-parties without first being de-identified, and there are no restrictions on how those predictions may be used. To its credit, Facebook claims its suicide predictions are never used for advertising. However, as the company becomes embroiled in one privacy scandal after another, it may be increasingly difficult for consumers to take the company at its word. Regardless, Facebook is one of many companies making mental health and suicide predictions. Without industry-wide scrutiny and stronger regulation, there will be ample opportunities for abuse.

3. Autonomy Risks

As described above, last year Facebook allegedly enabled advertisers to discriminate against minorities and people with disabilities by excluding them from receiving housing ads. As tech companies increasingly shape people's experiences online and in the real-world, they make decisions on their behalf, potentially depriving them of some degree of autonomy.

One side effect of suicide predictions is that people labeled high risk for suicide may be denied personal and professional opportunities, and in some cases, they may be deprived of civil liberties. The following sections describe how people labeled high risk for suicide may be deprived of opportunities to express themselves on Internet platforms and how their Fourth Amendment rights may be violated through warrantless searches based on opaque suicide predictions.

1. Censorship

Increasingly, platforms like YouTube, Twitter, and Facebook serve as 21st Century equivalents of the town square where people traditionally gathered to share ideas.[183] Internet platforms go to great lengths to moderate online conversations and maintain civility.[184] They have detailed

---

[182] Mason Marks, *Suicide prediction technology is revolutionary. It badly needs oversight*, WASH. POST (Dec. 20, 2018), https://www.washingtonpost.com/outlook/suicide-prediction-technology-is-revolutionary-it-badly-needs-oversight/2018/12/20/214d2532-fd6b-11e8-ad40-cdfd0e0dd65a_story.html?utm_term=.2f4c99f2a344.
[183] Zeynep Tufekci, *Twitter Has Officially Replaced the Town Square*, WIRED (Dec. 12, 2017), https://www.wired.com/story/twitter-has-officially-replaced-the-town-square/.
[184] Kate Klonick, *The New Governors: The People, Rules, and Processes Governing Online Speech*, 131 Harv. L. Rev. 1598 (2018).

community guidelines that govern what people can and cannot say, and users are routinely censored or banned for violating the rules.[185]

The New York Times recently described how Facebook's global speech rules are made: "Every other Tuesday morning, several dozen Facebook employees gather over breakfast to come up with the rules, hashing out what the site's two billion users should be allowed to say." Facebook distributes its speech guidelines to about 15,000 content moderators that it employs globally.[186] According to reports from some moderators, they have mere seconds in which to decide whether content is permissible or objectionable, which makes offloading some of the burden onto AI a necessity.

With over two billion users worldwide, Facebook's guidelines allow it exercise significant control over global speech. According to its community standards, moderators remove "content that encourages suicide or self-injury, including real-time depictions that might lead others to engage in similar behavior." However, these guidelines are applied inconsistently, and users have little recourse if Facebook removes their content.[187] Some users report having suicide notes removed from the platform while others report difficulty having them removed.[188]

In 2017, fourteen-year-old British teen Molly Russell killed herself.[189] In 2019, her father publicly claimed Instagram helped kill his daughter by failing to censor content that promotes and glorifies suicide.[190] In response to the story, British Secretary of State for Health Matt Hancock suggested Parliament could ban Internet platforms that fail to remove harmful content from their sites. Meanwhile, facing mounting pressure to improve the fairness of its content moderation, Facebook announced it would create an external board of independent experts to review its "most challenging content decisions."[191] Facebook promises the board will be composed of experts with experience in safety, privacy, and civil rights.[192]

The public health effects of censoring suicide-related speech are unknown. There is some evidence suggesting that increased media coverage of suicides promotes copycats and increases suicide rates.[193] However, it is unclear what effect censoring suicide-related speech on social media has on suicide rates.[194] Unlike the speech of news media, which is protected from government censorship by the First Amendment, the speech of social media users is not

---

185 Max Fisher, *Inside Facebook's Secret Rulebook for Global Political Speech,* NY TIMES (Dec. 27, 2018), https://www.nytimes.com/2018/12/27/world/facebook-moderators.html.
186 *Id.*
187 *See* Ariana Tobin et al., *Facebook's Uneven Enforcement of Hate Speech Rules Allows Vile Posts to Stay Up,* ProPublica (Dec. 28, 2017), https://www.propublica.org/article/facebook-enforcement-hate-speech-rules-mistakes.
188 *See e.g.* Deleting a suicide note? https://www.facebook.com/help/community/question/?id=1572559226321952.
189 Mile Wright and James Cook, *Sir Nick Clegg says Facebook has saved 'thousands' from suicide,* TELEGRAPH (Jan. 28, 2019), https://www.telegraph.co.uk/news/2019/01/28/sir-nick-clegg-says-facebook-has-saved-thousands-suicide/.
190 *Id.*
191 Draft Charter: An Oversight Board for Content Decisions, Facebook, https://fbnewsroomus.files.wordpress.com/2019/01/draft-charter-oversight-board-for-content-decisions-1.pdf.
192 *Id.*
193 Madelyn S. Gould, *Suicide and the Media,* 932 CLINICAL SCI. SUICIDE PREVENTION 200 (2001).
194 Thomas Ruder et al., *Suicide Announcement on Facebook,* 32 CRISIS: J. CRISIS INTERVENTION SUICIDE PREVENTION 280 (2011).

protected because Internet platforms and their content moderators are not government entities. Nevertheless, there may be public health arguments for ensuring freedom of expression for users of online platforms.

Though it is possible that uncensored suicide-related speech could inspire copycats, it is equally plausible that stifling public discussion of suicide contributes to its taboo nature and inhibits people from seeking and receiving needed help and support. Somewhat surprisingly, Facebook does not censor suicide-related expression when users live-stream their suicide attempts. Its rationale is that leaving the stream running “until the point of no return” maximizes the chance that viewers of the stream can send for help. The problem is Facebook makes these decisions unilaterally, censoring some instances of suicide-related speech, but not others, and its decisions are not transparent or evidence-based.

2. Warrantless Searches

As AI-based suicide prediction tools proliferate, they will play an increasing role in police and doctors’ decisions to involuntarily hospitalize people for treatment or medical observation. Civil commitment is an intervention that strips people of liberty and autonomy, and it is not without risks.[195] Nevertheless, it is permitted by state laws when individuals are deemed a risk to themselves or others.[196] If a person is deemed high-risk by social suicide prediction tools, prompting police officers to respond to that person’s home, and the person does not answer the door, then police could enter the home without first obtaining a search warrant.

In the US, the Fourth Amendment protects people and their homes from warrantless searches.[197] However, under exigent circumstances doctrine, police may enter homes without warrants if they reasonably believe entry is necessary to prevent physical harm. Stopping an imminent suicide attempt clearly falls within this exception. However, it may be unreasonable to rely on opaque AI-generated suicide predictions to circumvent Fourth Amendment protections when no information regarding their accuracy is publicly available. As described above, Facebook and Crisis Text Line make suicide predictions based on internal data rather than data from real suicides. We don’t know how accurate their predictions are, what criteria they use to decide when law enforcement should be contacted, or what information they provide to police. Exceptions to the warrant requirement should not be made based on such paltry information.

III. A Policy Framework for Regulating AI-Based Suicide Prediction

Part I of this article described how companies use AI to infer suicide risk from medical records and consumer behavior. Part II explained the risks to people’s safety, privacy, and autonomy. This part proposes a policy framework for minimizing those risks. The recommendations are inspired by laws that govern medical practice and biomedical research, such as HIPAA and the Federal Common Rule, as well as new privacy laws introduced in California and the European

---

[195] [Stripped of liberty and autonomy]
[196] [State civil commitment law]
[197] Ken Wallentine, *Should I Stay or Should I Go- If you respond to a call involving a suicidal person who’s not endangering anyone else, it might be best to not intervene*, POLICE MAGAZINE (Oct. 16, 2017), http://www.policemag.com/channel/patrol/articles/2017/10/should-i-stay-or-should-i-go.aspx.

Union. Though these recommendations will not eliminate all risks, they are a good starting point. Companies that make suicide predictions can use them as a template for implementing self-imposed standards for suicide prediction. The framework can also serve as a foundation for laws to regulate suicide predictions in the US and internationally.

A. Suicide prediction research should be approved by independent IRBs, and ongoing suicide prediction programs should be monitored for safety and efficacy by independent data monitoring committees.

In the US, drugs and medical devices are tested for safety and efficacy through clinical trials conducted with FDA oversight. Before testing begins, trial protocols are reviewed and approved by IRBs at the institutions conducting the research. In some cases, after clinical trials commence, their progress is observed by data monitoring committees (DMCs). DMCs are composed of people with relevant expertise who conduct ongoing review of clinical trial data as it is generated.[198] They make recommendations to trial sponsors regarding “the continuing safety of trial subjects” and the “continuing validity and scientific merit of the trial.”[199] If a DMC determines a trial is no longer safe or scientifically valid, it may recommend the trial be stopped.[200]

To ensure the safety of patients and consumers, suicide prediction research should be reviewed and approved by independent IRBs, and once implemented, suicide prediction programs should be monitored for safety and efficacy by independent DMCs. Those IRBs and DMCs must be truly independent, having no financial connections to the companies making predictions and no stake in the outcomes of their research. As discussed previously, Facebook has an internal ethics review board. However, because this body is composed of Facebook employees and its review of Facebook’s research is optional, it is not an effective safeguard against the risks posed by social suicide predictions. Instead, Facebook should use an independent review board comparable to the oversight board it recently proposed to review content decisions.[201] Social suicide predictors are essentially conducting large unregulated clinical trials in which suicide predictions are made and interventions are initiated. Those predictions and interventions affect real people’s lives and may result in serious injury or death. However, there is no ongoing, independent review of their methods and outcomes. Independent IRB and DMC review of predictions would provide needed oversight.

To be fair, not all clinical trials have DMCs. However, the FDA recommends researchers consider using DMCs for trials involving potential fragile or vulnerable populations.[202] According to one study on DMCs “Most psychiatric patients meet FDA standards for a vulnerable or high-risk population, so based on the above recommendations most studies

198

199

200

201 *See* Facebook *supra* note.

202 Food and Drug Administration, *Establishment and Operation of Clinical Trial Data Monitoring Committees for Clinical Trial Sponsors* (last updated Jul. 12, 2018), https://www.fda.gov/RegulatoryInformation/Guidances/ucm127069.htm.

involving psychiatric patients should have DMCs."[203] DMCs are usually composed of statisticians, researchers having expertise in other relevant fields, patient representatives, and if a study involves vulnerable populations, its DMC may include members of those groups or their relatives. A DMC for evaluating the safety and efficacy of suicide predictions should be composed of statisticians, psychiatrists, psychologists, bioethicists, privacy experts, people who have suicidal thoughts or people who have attempted suicide, and members of vulnerable groups who are disproportionately affected by suicide or suicide prediction methods such as veterans, Native Americans, and members of the LGBT community.

Because medical suicide prediction is conducted by hospitals and healthcare systems, it must comply with federal research regulations and general principles of medical ethics by promoting patient autonomy, justice, beneficence, and non-malfeasance.[204] However, unlike medical suicide prediction, most social suicide prediction is unregulated and subject to none of these requirements. Thus, there is currently no way to evaluate whether social suicide prediction is safe and effective unless this framework or similar standards are implemented.

There are of course many details to work out. Who should convene and chair these IRBs and DMCs, and who should pay for them? Would social suicide predictors be bound by their recommendations, and who should enforce compliance? Nevertheless, as tech companies start taking on health-related roles that were traditionally reserved for doctors and public health agencies, there must be mechanisms in place to oversee their operations and promote public safety.

B. Suicide prediction methods should be transparent and made available to consumers and external suicide researchers.

Suicide is a national public health problem, and thousands of lives are put at risk when suicide interventions are made. Yet social suicide predictors maintain their algorithms as proprietary trade secrets. Instead, consumers should demand transparency, and suicide predictors should be required to share their methods with consumers and suicide researchers in the greater scientific community.

Greater transparency and information sharing would help ensure that suicide prediction algorithms are safe, and it would allow outside researchers, such as medical suicide predictors, to benefit from knowledge gained through social suicide predictions. Publicly shared algorithms could be scrutinized by computer scientists, privacy experts, and mental health professionals to ensure that data is stored securely and is not transferred to data brokers and advertisers.

Facebook and Crisis Text Line may have intellectual property-related reasons for keeping their algorithms secret. For example, Facebook's algorithms may contain proprietary technology and share features with the company's advertising systems. Making them public could decrease the company's competitive advantage in the advertising space and expose its systems to increased scrutiny. However, these concerns must be weighed against the public health risks of keeping the

[203] Julia Y. Lin and Ying Lu, *Establishing a data monitoring committee for clinical trials,* 26 SHANGHAI ARCHIVES PSYCHIATRY 26 (2014).
[204] Tom L. Beauchamp, *Methods and principles in biomedical ethics,* 29 J. MED. ETHICS 269 (2003).

algorithms secret and untested. Unless social suicide predictors can establish that their methods are safe, we can't be sure they don't contribute to the problems they are intended to alleviate.

If suicide prediction methods were publicly disclosed, then members of the public would be better informed, and they could make fully-informed decisions to use products and services that make suicide and mental health-related predictions. There are non-trivial costs associated with wellness checks, and taxpayers must foot the bill for the use of emergency services. Therefore, the public deserves to know whether those interventions are safe and effective. Suicide predictions and interventions may disproportionately affect vulnerable populations exacerbating existing societal inequalities and creating negative externalities that are borne by those groups, and by society, but not by social suicide predictors.

Transparency is a hallmark of the open source software community, and multinational tech companies are increasingly giving their patented technology to the public. In 2018, Microsoft made 60,000 patents "open source," allowing anyone to use them and potentially forgoing billions of dollars in royalties.[205] In a recent interview with 60 Minutes, Tesla CEO Elon Musk said "I'm not sure if you know it, but we open sourced our patents, so anyone who wants to use our patents can use 'em for free."[206] Musk said he would be happy if another company used Tesla's technology to make a better electric car, even if it put Tesla out of business, because it would be good for the environment.[207] Considering the devastating effects of suicide on families and communities, and the negative externalities that arise from inaccurate predictions, Facebook should take a similar stance with respect to its suicide prediction methods.

C. Suicide prediction programs should be opt-in only and provide patients and consumers with clear methods to opt-out and delete their information.

Today's consumers are tracked on an unprecedented scale, often without their knowledge or consent.[208] Data is sometimes described as the "new gold" or the "new oil." Consumer information is mined and widely bought and sold. Social media platforms are a major source of the data, and sensitive information extracted from these sites can be sold to data brokers, employers, lenders, and insurance companies.[209] The information is often health-related revealing deeply personal information about people's medical and psychological traits.[210]

---

[205] Jason Evangelho, *60,000 Patents, Proving It Really Love Linux*, FORBES (Oct. 11, 2018), https://www.forbes.com/sites/jasonevangelho/2018/10/11/microsoft-just-open-sourced-60000-patents-proving-it-really-does-love-linux/#4bcf11463807.
[206] Leslie Stahl, *Tesla CEO Elon Musk: The 60 Minutes Interview*, 60 MINUTES (Dec. 9, 2018), https://www.cbsnews.com/news/tesla-ceo-elon-musk-the-2018-60-minutes-interview/.
[207] *Id.*
[208] *See e.g.* Janus Kopfstein, *Verizon Is Still Tracking Customers Across the Web Without Consent*, MOTHERBOARD (Mar. 9, 2016), https://motherboard.vice.com/en_us/article/wnxdwy/verizon-supercookie-tracking-loophole.
[209] Marks *supra* note.
[210] *Id.*

When signing up for online services, consumers must agree to "click wrap" agreements that are required to use most online platforms.[211] The only way to decline is to forego using the services. For instance, Facebook users cannot opt out of having their data mined for the company's suicide prediction program. The only way to opt-out is to refuse to sign up for Facebook, stop using site, or delete one's account. Users of other services such as Crisis Text Line and Operation Zero may consent to having their information collected without realizing what they are agreeing to. For example, Crisis Text Line users engage the service by sending texts from their smartphones. They may use the service without ever visiting its website or reading terms of service. People presumably turn to crisis support services in times of great stress. Even if they do read the terms, they may not be capable of providing informed consent while under duress. Moreover, then percent of Crisis Text Line's users are under thirteen, and they may not fully understand the risks of having their data mined by the service. Instead of hiding their data collection practices in fine print, Crisis Text Line should ask users in plain language whether they wish to opt-in to its suicide prediction program when they first contact the service. It should also inform users that their data may be transferred to the company's for-profit spinoff Loris.AI.

According to Facebook's Emily Cain "By using Facebook, you are opting into having your posts, comments, and videos (including FB Live) scanned for possible suicide risk."[212] However, the word "suicide" does not appear in Facebook's data policy.[213] Instead, the policy states "we use data we have to . . . <u>detect when someone needs help</u>."[214] This statement is vague, and it is buried in the middle of the data policy that is over 2,000 words long.[215] Unless users read through half the document and click on an embedded link, which takes them to an article about suicide prevention tools, they would be unaware that Facebook makes suicide predictions. Neither the data policy nor the page it links to contain information about wellness checks or scanning of user videos and live streams.[216] The lack of information casts doubt on Facebook's claim that user's knowingly opt-in to having all their content scanned for suicide risk. Moreover, research suggests most users don't read privacy or data policies. One study found that over 90% of US consumers agree to legal terms of service without reading them.[217] The rate is 97% for Americans between the ages of 18 – 34.[218]

When users sign up for services that make suicide predictions, they should be given the option to opt-in to suicide prediction services. They should be warned in prominent, easy to read language that their data will be used to calculate a suicide risk score, and if the score is high enough, police may be sent to their homes, which could result in warrantless searches of their homes and involuntary hospitalization and forced medication. Consumers deserve to know this information because without it, they cannot make an informed decision to assume the associated risks.

---

[211] *See e.g.* Eriq Gardner, *Canadian Supreme Court Doesn't Like Facebook Clickwrap in Big Privacy Suit,* Hollywood Reporter (Jun. 23, 2017), https://www.hollywoodreporter.com/thr-esq/canadian-supreme-court-doesnt-like-facebook-clickwrap-big-privacy-case-1016277.
[212] Goggin *supra* note 78.
[213] Data Policy, Facebook, https://www.facebook.com/policy.php.
[214] *Id.*
[215] *Id.*
[216] *Id.*
[217] Caroline Cakebread, *You're not alone, no one reads terms of service agreements,* BUS. INSIDER (Nov. 15, 2017) https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11
[218] *Id.*

If companies update their data or privacy policies, then users should be asked to reaffirm their choice to have suicide-related inferences and interventions made on their behalf. Such a program would respect people's autonomy. When the VA created the Durkheim Project, it decided to require veterans to opt-in instead of enrolling them automatically in the program. Similar opt-in systems should be the gold standard for medical and social suicide predictions.

D. Social suicide predictors should treat their predictions as sensitive health data and protect them through compliance with HIPAA-like standards.

Facebook and other suicide predictors claim their predictions do not constitute medical information and that they are not acting as healthcare providers.[219] However, by performing suicide risk assessments on their users, these companies are taking on roles historically reserved for healthcare providers and public health agencies. The predictions they make constitute sensitive health-related information that is no less sensitive than the data contained in medical records.

In the past, suicidal thoughts and risk predictions might have been shared only with doctors, psychotherapists, family members, and spiritual advisors. However, due to the proliferation of AI and big data, companies can now infer this information, and the act of making those inferences closely parallels the diagnostic process performed by doctors and other healthcare providers. Consider the VA's Durkheim Project, which analyzed veterans' social media activity. Because it occurred within the VA health system, it was subject to health laws, biomedical research regulations, and general principles of medical ethics. And because these predictions were made by healthcare providers, they were considered part of the practice of medicine. However, when Facebook makes predictions using nearly identical technology, it is unregulated, and few people recognize the process as comparable to medical practice.

How can the same technology be heavily regulated in one context and almost completely unregulated in another? The double-standard is due to outdated health laws that have not kept up with rapid changes in technology. Nevertheless, the health and privacy concerns that gave rise to current laws such as HIPAA are no less applicable to social suicide predictions than to traditional health data. For instance, the legislative history of HIPAA reveals it was drafted to address concerns that exposing people's health data to third-parties would lead to exploitation, and those concerns apply equally well to social suicide prediction data. The fact that most social suicide predictors are not covered entities under HIPAA does not make suicide predictions non-health data. Suicide predictions are health data regardless of their source, and suicide prediction programs are health screening programs, plain and simple. To protect user data, social suicide predictors should comply with HIPAA privacy and security requirements, even if compliance is not required by law.

"The HIPAA Privacy Rule establishes national standards to protect individuals' medical records and other personal health information." "The Security Rule operationalizes the protections contained in the Privacy Rule by addressing the technical and non-technical safeguards that

[219] *See e.g.* Kaste *supra* note 7.

organizations called "covered entities" must put in place to secure individuals' "electronic protected health information" (e-PHI)."

E. Suicide prediction-related data should not be shared with third parties or used for advertising.

As described in Part II, sharing suicide predictions with third-parties such as data brokers and advertisers can promote stigmatization and exploitation of consumers. Targeted ads may be designed to exploit depressed or suicidal people's vulnerabilities and deny them access to resources such as housing or employment based on their suicide risk scores. Laws such as the Americans with Disabilities Act prohibit some forms of discrimination. However, they often put the burden on consumers to identify instances of discrimination, which can be challenging when algorithms are run without transparency.

A new wave of privacy laws will encourage transparency. For instance, California's Consumer Protection Act requires each business that sells consumer information to disclose "the category or categories of consumers' personal information it has sold, or if the business has not sold consumer's personal information, it shall disclose that fact." Section 1798.120(a) gives consumers the right to opt-out of having their personal information sold. Businesses must notify consumers of this right.

Companies that make suicide predictions should be barred from using those predictions for advertising, profiling, or other uses that may exploit consumers.

F. "Soft touch" suicide interventions should be preferred over "firm hand" interventions.

Police forces in some cities train officers to serve on a specialized crisis intervention team (CIT).[220] The CIT model was developed in response to situations in which officers used deadly force while responding to mental health-related calls. An organization called CIT International "aspires to be a leader in promoting safe and humane responses to those experiencing a mental health crisis." Using a CIT model, the Memphis Police Department teaches officers that people with mental illness are more likely to be victims of violent crimes than perpetrators.[221] Officers with the San Antonio Police Department's mental health unit wear plain clothes and drive unmarked cars.[222] When they approach people with mental illnesses, they use speech, body language, and other techniques intended to de-escalate situations. Instead of referring to people as suspects, they refer to them as consumers and assure them they are not in trouble.

A CIT approach may have prevented the deaths of Jason Harrison, Anthony Hill, John Albers, and others with mental illnesses who have been shot and killed during police interventions.

---

[220] PBS News Hour, *How Memphis has changed the way police respond to mental health crises,* YOUTUBE, https://www.youtube.com/watch?v=y99kODtyVhk.
[221] *Id.*
[222] ABC News, *Meet Police Officers Trained to Respond to Mental Illness Calls,* YOUTUBE (Oct. 1, 2015), https://www.youtube.com/watch?v=sggp-mpXL4c; HopeForChildren1, *Law Enforcement and People with Mental Illness- Part 1,* YOUTUBE (Jun. 29, 2016), https://www.youtube.com/watch?v=U6Pw_dC05D8.

However, according to the National Alliance on Mental Illness, “only 15 – 20% of law enforcement agencies in the country have CIT programs.” Survey from Police Executive Research Forum “new recruits typically spend 60 hours learning to handle a gun compared to 8 hours learning strategies for handling the mentally ill.”

In many countries, police officers do not carry firearms: “Britain, Ireland, Norway, Iceland and New Zealand,” Scotland?[223] If entire European countries can maintain police forces where most officers to not carry firearms, then police in the US should be able to perform wellness checks on people with mental illnesses without weapons draw.

If social suicide predictors continue to initiate wellness checks in communities around the world, they may have a moral responsibility to contact only law enforcement agencies that use soft-touch or CIT approaches to suicide intervention. Social suicide predictors should invest in creating CIT teams and training them to favor soft-touch crisis interventions.

CONCLUSION

Accurately predicting suicide is an important goal. However, while healthcare providers and medical researchers take a cautious approach to implementing suicide predictions, tech companies are leading an aggressive charge to implement the technology. When those companies leverage consumer data to make suicide predictions, they take on roles that were once reserved for healthcare providers. Current health laws do not recognize their suicide predictions as protected health information. Nevertheless, without greater oversight, transparency, and accountability, there is potential for serious harm. The tools intended to protect suicidal people may promote exploitation, increase the risk of death, and inadvertently marginalize the groups they are meant to serve. Clear standards are needed to ensure fairness, safety, and effectiveness.

[223] Rick Noack, *5 countries where most police officers do not carry firearms- and it works well*, Wash. Post (Jul. 8, 2016), https://www.washingtonpost.com/news/worldviews/wp/2015/02/18/5-countries-where-police-officers-do-not-carry-firearms-and-it-works-well/?noredirect=on&utm_term=.c0b059627855.