Journal of Law and the Biosciences, 1–11
https://doi.org/10.1093/jlb/lsab021
Original Article

Legal, ethical, and wider implications of suicide risk detection systems in social media platforms

Karen L. Celedonia[1,*], Marcelo Corrales Compagnucci[2], Timo Minssen[2] and Michael Lowery Wilson[3]

[1]Injury Epidemiology and Prevention, Turku Brain Injury Centre, University of Turku and Turku University Hospital, Turku, Finland
[2]Center for Advanced Studies in Biomedical Innovation Law (CeBIL), University of Copenhagen, Copenhagen, Denmark
[3]Heidelberg Institute of Global Health (HIGH), University of Heidelberg, Heidelberg, Germany

*Corresponding author. E-mail: klceledonia@gmail.com

ABSTRACT

Suicide remains a problem of public health importance worldwide. Cognizant of the emerging links between social media use and suicide, social media platforms, such as Facebook, have developed automated algorithms to detect suicidal behavior. While seemingly a well-intentioned adjunct to public health, there are several ethical and legal concerns to this approach. For example, the role of consent to use individual data in this manner has only been given cursory attention. Social media users may not even be aware that their social media posts, movements, and Internet searches are being analyzed by non-health professionals, who have the decision-making ability to involve law enforcement upon suspicion of potential self-harm. Failure to obtain such consent presents privacy risks and can lead to exposure and wider potential harms. We argue that Facebook's practices in this area should be subject to well-established protocols.[1] These should resemble those utilized in the field of human subjects research, which upholds standardized, agreed-upon, and well-recognized ethical practices based on generations of precedent. Prior to collecting sensitive data from social media users, an ethical review process should be carried out. The fiduciary framework seems to resonate with the emergent roles and obligations of social media platforms to accept more responsibility for the content being shared.

KEYWORDS: suicide risk detection, social media platforms, algorithms, AI, ethics, privacy, consent, legal implications

I. INTRODUCTION

Someone dies by suicide somewhere in the world every 40 seconds. Suicide has the unenviable distinction of being the second leading cause of death for individuals aged 15–29 years[1] of age worldwide.[2] Suicide, as well as non-suicidal self-harm, are global problems of public health importance.[3]

Much has been written about the implications of media influences on suicidal behavior,[4] even as far back as the Great Depression[5] and World War I.[6] In recent decades, there has been a revived interest in media influences on self-harm. The renewed discussion has focused on the subject of social media and the depth to which it transforms human behavior.[7] Specifically, social media exposure may have a compounding effect on suicidal behavior.[8] Yet at the same time, it also represents a potential avenue for suicide prevention.[9] Social media platforms have been accused of not doing enough to prevent suicidal behavior.[10] But social media platforms' response to these accusations, which has included bolstering suicide detection mechanisms on the platforms, has been met with criticism as well, with the primary point of contention being the implementation of an intervention that was seemingly hastily devised without proper consideration of the social and cultural context in which an individual experiencing suicidal behavior is embedded. The main question within this context has been who should be engaged in the discovery and prevention of suicidal behavior on social media platforms and to what extent? Facebook in particular has been the subject of considerable debate in its ongoing use of Artificial Intelligence (AI) and algorithms in an attempt to detect suicidal thoughts and actions on its platform.[11,12] These discussions fall mainly within

1 This paper could consider developments until January 2021.

2 Norberto Nuno Gomes de Andrade et al., *Ethics and Artificial Intelligence: Suicide Prevention on Facebook*. 31 Philos. Technol. 669–684 (Dec. 2018), https://doi.org/10.1007/s13347-018-0336-0.

3 Florian Arendt, *Suicide on Instagram—Content Analysis of a German Suicide-Related Hashtag*, 40 Crisis 36–41 (Jan. 2019).

4 Darren Baker and Sarah Fortune, *Understanding Self-Harm and Suicide Websites: A Qualitative Interview Study of Young Adult Website Users,* 29 Crisis 118–122 (2008), Erin L. Belfort and Lindsey Miller, *Relationship Between Adolescent Suicidality, Self-Injury, and Media Habits,* 27 Child and Adolescent Psychiatric Clinics of North America 159–169 (2018) and Qijin Cheng et al., *Media Effects on Suicide Methods: A Case Study on Hong Kong 1998–2005*, 12 PloS One e0175580 (Mar. 2017).

5 Steven Stack, *The Effect of the Media on Suicide: The Great Depression*, 22 Suicide & Life-Threatening Behavior 255–267 (1992).

6 Steven Stack, *Suicide: Media Impacts in War and Peace, 1910–1920*, 18 Suicide & Life-Threatening Behavior 342–357 (1988).

7 Audrey Poh Choo Cheak et al., Online Social Networking Addiction: Exploring its Relationship With Social Networking Dependency and Mood Modification Among Undergraduates in Malaysia, in International Conference on Management, Economics and Finance, Sarawak, Malaysia (2012).

8 D. David et al., *Social Media and Suicide: a Public Health Perspective*, 102 American Journal of Public Health S195–S200 (2012).

9 Jo Robinson et al., *Social Media and Suicide Prevention: a Systematic Review,* 10 Early Intervention in Psychiatry 103–121 (Apr. 2016).

10 Caroline Warnock, *Ronnie McNutt's Friend Speaks Out About Facebook Suicide Video*. Heavy (2020). Retrieved from https://heavy.com/news/2020/09/ronnie-mcnutt-facebook-video.

11 Ian Barnett and John Torous, *Ethics, Transparency, and Public Health at the Intersection of Innovation and Facebook's Suicide Prevention Efforts*, Annals of Internal Medicine (2019), https://www.acpjournals.org/doi/abs/10.7326/M19-0366?journalCode=aim.

12 Tineke Broer, *Technology for Our Future? Exploring the Duty to Report and Processes of Subjectification Relating to Digitalized Suicide Prevention,* 11 Information 170 (2020), doi: 10.3390/info11030170.

the context of a perceived public health imperative to protect those with demonstrated behavioral manifestations of suicidal behavior, and the ethical and legal boundaries for automating the detection of suicidal behaviors in public virtual spaces. Where these discussions fall short however is in considering how intricately nuanced societal and cultural factors come into play with regard to mental health stigma in the context of suicide and the role of informed consent in predictive analytics.

Recent research has pointed towards links between the global rise in suicidal behavior among adolescents with the increase in social media use.[13] It has been argued that social media platforms encourage and sustain anti-social behaviors such as envy and pathological narcissism[14] via the so called 'network effect'.[15] Emerging literature also suggests that social media use is correlated with increased risk of depression and anxiety disorders, both of which are risk factors for suicidal behavior.[16] Furthermore, it is not only general social media use itself that potentially poses harm to users, but also the opportunity social media presents to propagate negative social behaviors such as cyber bullying. A large amount of cyber bullying takes place on social media platforms, and bullying victimization is another well-established risk factor for suicidal behavior.[17] The negative effects of social media use on users' mental health could therefore at least partially explain the link between increased suicidal behavior and social media use.

In light of the growing body of research connecting social media use to a host of mental health problems, including suicidal behavior, and rapid technological advances making suicide prediction from clinical notes a useful tool in augmenting therapeutic decision-making,[18] a suicide detection algorithm on social media platforms like Facebook might arguably seem akin to a teacher identifying behavior in students that is indicative of underlying mental health conditions requiring treatment or detecting signs of potential child abuse and neglect. The teacher or other child care worker is duty-bound to report such issues to the proper authorities so that the necessary interventions are performed to help and protect the child. However, what distinguishes Facebook's suicide detection algorithm from these scenarios is that by collecting and acting on user data, they are moving beyond the role of mandated reporter to arguably behaving like mental health care providers. Would it be ethically sound for a teacher, who is not a mental health care professional, to provide mental health treatment or recommendations to a student with mental illness? To further compound this ethical dilemma, suicidal behavior, within the context of social media prediction, is fundamentally a socio-behavioral and a socially stigmatized problem occurring within

13 David D. Luxton et al., *Social Media and Suicide: A Public Health Perspective*, American Public Health Association (APHA) publications (Apr 2012) and Hong-Hee Won et al., *Predicting National Suicide Numbers with Social Media Data*, 8 PLoS One e61809 (Apr. 2013).

14 Siva Vaidhyanathan, Antisocial Media: How Facebook Disconnects Us and Undermines Democracy, Oxford University Press (2018) and Jan Fox, *An Unlikeable Truth: Social Media Like Buttons are Designed to be Addictive, They're Impacting Our Ability to Think Rationally*, 47 Index on Censorship 11–13 (2018).

15 Wenlin Liu et al., Network Theory, Wiley Online Library, pages 1–12 (Mar. 2017).

16 Betul Keles et al., *A Systematic Review: the Influence of Social Media on Depression, Anxiety and Psychological Distress in Adolescents*, 25 International Journal of Adolescence and Youth 79–93 (2020).

17 David D. Luxton et al., *Social Media and Suicide: A Public Health Perspective*, 102 American Journal of Public Health S195–S200 (May 1, 2012), https://doi.org/10.2105/AJPH.2011.300608.

18 Chris Poulin et al., *Predicting the risk of suicide by analyzing the text of clinical notes*. PLoS One, 24489669. Retrieved from https://pubmed.ncbi.nlm.nih.gov/24489669.

(and outside of) a virtualized social environment. Thus, any data that are fed into a suicide prediction algorithm remain subject to the underlying biases inherent to these settings, and the extent to which said settings can in and of themselves modify behavior. These biases can bleed into the outputs of algorithmic prediction with very uncertain consequences.[19] So, with Facebook's suicide detection algorithm myriad ethical and regulatory concerns exist and remain disregarded. We argue that the mainstreaming of algorithmic suicide risk detection may represent unmitigated societal issues where the legal, ethical, cultural and regulatory issues are not duly considered.

II. LEGAL CONSIDERATIONS

In the USA, information about individuals' health is protected by the Health Insurance Portability and Accountability Act (HIPAA),[20] which requires specific privacy protections, including encryption and sharing restrictions, when handling health records. These rules, however, only apply to organizations providing healthcare services and plans such as hospitals and insurance companies.[21] Since Facebook is not a health services provider, one of the first issues to address is whether social media data generated by individual users would count as personal health information and whether it should be protected by HIPAA. Suicide prediction in healthcare settings, or medical suicide prediction, as referred to by Marks, falls under HIPAA rules. This is because AI is applied against healthcare records in order to analyze patient records for patterns to predict the future probability of suicidal behavior. However, social media platforms such as Facebook's use of AI to predict suicide or 'social suicide prediction' generally fall outside the scope of HIPAA and are thus not covered under HIPAA rules.[22] In social media prediction, algorithms examine digital traces of human behavior, which provide intricate clues about an individual's health status. Some of this information is given intentionally, such as a text message about the course of one's day to another user on the platform. Other information may not be given intentionally, such as a location or conducting searches using health-related terms. This fact raises the fundamental question of whether Facebook is beholden to the provisions enshrined in HIPAA legislation. Our main consideration here is that algorithms[23] could potentially parse various forms of personal health information such as disclosure of medication use, doctor visits,[24] or even reveal details about the health of persons intimately connected

19 Trishan Panch et al., *Artificial Intelligence and Algorithmic Bias: Implications for Health Systems*. 9 J. Glob. Health. doi:10.7189/jogh.09.020318.

20 Health Insurance Portability and Accountability Act (HIPAA); Kennedy–Kassebaum Act, or Kassebaum–Kennedy Act).

21 Karen Celedonia et al., *Community-Based Health Care Providers as Research Subject Recruitment Gatekeepers: Ethical and Legal Issues in A Real-World Case Example, Research Ethics* (2020), https://doi.org/10.1177/1747016120980560 (last visited Jan. 18, 2021).

22 Mason Marks, *Artificial Intelligence for Suicide Prediction*, Bill of Health (Nov. 6, 2018) https://blog.petrieflom.law.harvard.edu/2018/11/06/artificial-intelligence-for-suicide-prediction/ (last visited Jan. 18, 2021).

23 Ugo Pagallo et al., *The Rise of Robotics & AI: Technological Advances & Normative Dilemmas*, in Robotics, AI and the Future of Law 2, 6, 10 (Marcelo Corrales, Mark Fenwick and Nikolaus Forgó, eds., Springer, 2018).

24 Erik P.M. Vermeulen et al., *Business and Regulatory Responses to Artificial Intelligence: Dynamic Regulation, Innovation Ecosystems and the Strategic Management of Disruptive Technology*, in Robotics, AI and the Future of Law 83–84 (Marcelo Corrales, Mark Fenwick and Nikolaus Forgó, eds., Springer 2018).

to users such as a spouse or children who may not themselves be social media users or consent to have their information used in the manner proposed by the social media giant. Therefore, transparency with regard to research activity and the necessity of informed consent from research participants is of paramount importance.[25]

Users of Facebook, in general, appear to be unaware of the platform's suicide risk detection strategy.[26] If this is indeed the case then it would suggest that users are not fully knowledgeable of what they are consenting to under Facebook's terms and conditions regarding the use of their data. Consent should be freely given and information regarding the suicide prevention goals and how the algorithm works should be duly provided to the participants and specific to the project. Failure to obtain such consent presents privacy risks and can lead to exposure and harm individuals. This is the main reason why Facebook's suicide algorithm is banned in the European Union (EU).[27] Stricter rules due to the General Data Protection Regulation (GDPR)[28] requires users to provide websites specific consent[29] to collect sensitive data such as that pertaining to someone's mental health.[30]

Facebook is browsing virtually every post on the platform in an attempt to detect signs of potential suicide risk. Then, Facebook passes the information along to a law enforcement agency for wellness checks.[31] However, as a private media company that derives significant income from advertising and shaping public opinion,[32] it has none of the ethical oversight or privacy protections in place regarding the collection and synthesis of this data. Following a string of data leak scandals,[33] it is doubted that Facebook can be in the position to protect users' sensitive data. Using a medical

25 Marcelo Corrales, *'Protecting Patients' Rights in Clinical Trial Scenarios: The 'Bee Metaphor' and the Symbiotic Relationship'*, in An Information Law for the 21st Century 5–13 (Maria Botties, ed., Nomiki Bibliothiki 2011).

26 Christopher Burr et al., *Digital Psychiatry: Ethical Risks and Opportunities for Public Health and Well-Being* (Oct. 30, 2019), https://ssrn.com/abstract=3477978 or http://dx.doi.org/10.2139/ssrn.3477978 (last visited Jan. 18, 2021).

27 Simone Osborne, *EU Rules BLOCK Facebook from Introducing a Tool to Stop Suicides Over Data Protection Breach*, Express (Nov. 29, 2017), https://www.express.co.uk/news/uk/885763/european-union-data-protection-rules-block-Facebook-suicide-prevention-tool (last visited Jan. 25, 2021).

28 Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

29 According to Recital 32 of the GDPR 'consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data . . . ', 157

30 Benjamin Goggin, *Inside Facebook's Suicide Algorithm: Here's How the Company Uses Artificial Intelligence to Predict Your Mental State From Your Posts*, https://www.businessinsider.com/facebook-is-using-ai-to-try-to-predict-if-youre-suicidal-2018-12?r=US&IR=T (last visited Aug. 3, 2020).

31 *Id.*

32 Vedava Baraković, *Facebook Revolutions: The Case of Bosnia and Herzegovina*, 1 Acta Universitatis Sapientiae. Social Analysis 194–205 (2011) and Milad Dehghani and Mustafa Tumer, *A Research on Effectiveness of Facebook Advertising on Enhancing Purchase Intention of Consumers*, 49 Computers in Human Behavior 597–600 (2015).

33 BBC News, *Facebook 'to be Fined $5bn Over Cambridge Analytica scandal'*, https://www.bbc.com/news/world-us-canada-48972327 (last visited Aug. 3, 2020); Laurence Dodds, *Facebook Was Repeatedly Warned of Security Flaws That Led to Biggest Data Breach in its History*, The Telegraph Feb. 9, 2020, https://www.telegraph.co.uk/technology/2020/02/09/facebook-repeatedly-warned-security-flaw-led-biggest-data-breach/ (last visited Aug. 3, 2020).

example, patients go to a healthcare provider based on its track record and institutional trust. The role of physicians is very important as they can take on a fiduciary obligation acting as a faithful trustee.[34] Therefore, it is useful to think of Facebook in the context of negligence and fiduciary duties taking into account they lack the same institutional trust and privacy standards as healthcare providers.

Like physicians, regulators could impose fiduciary duties on social media companies and its suicide prediction algorithm. This is a legal duty that would require companies like Facebook to act in their users' benefit and will ensure that there is no conflict of interest.[35] Balkin has suggested treating social media platforms as 'information fiduciaries'. This concept would rebalance the relationship between users and the online platform that accumulates, analyzes, and sells their personal data for profit. This fiduciary includes two basic duties: the 'duty of care' and the 'duty of loyalty'. In the first duty, fiduciaries must act competently and diligently. In the second duty, fiduciaries must act in the best interest of the beneficiary and make sure there are no actual or potential conflict of interests that might harm their clients. In addition, another important duty to mention in this context is the 'duty of confidentiality', which is also imposed on physicians. In practical legal terms, this means that all patient information must not be disclosed without the consent of the patient. Confidentiality is fundamental to the preservation of trust between doctors and their patients. The aim is to make patients feel secure and comfortable enough providing personal and sensitive information as this plays a crucial role in medical treatment.[36] The main exception to this duty of confidentiality is if the patient is believed to be a risk to themselves or others, in which case the physician becomes duty bound to report using the appropriate legal means.[37] An information fiduciary is, therefore, a person or company who takes special obligations of loyalty, trustworthiness and care towards the information of others.[38]

The fiduciary 'duty of care' might require Facebook to prove that its prediction algorithms have undergone an adequate safety assessment and efficacy testing by a competent third party. The 'duty of confidentiality' and 'duty of loyalty' could require Facebook to protect users' sensitive data, including refraining from putting the company's interests in front of profits.[39] In a recent interview between Facebook's CEO

34 Marcelo Corrales Compagnucci, Big Data, Databases and 'Ownership' Rights in the Cloud 289 (Springer 2020).

35 Mason Marks, *Suicide Prediction is Revolutionary, It Badly Needs Oversight. Should we Trust Facebook to Dispatch Police to the Homes of Distraught Users?* (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 (last visited Jan. 18, 2021).

36 Patient confidentiality is, however, not absolute. Exceptions are made for legitimate and public interest purposes such as in some circumstances which could risk lives or seriously harm other individuals. See, Anne Morris and Michael Jones, Blackstone's Status on Medical Law, 414 (OxfordUniversity Press, 10th ed, 2019), K. Blightman et al., *Patient Confidentiality: When can a Breach be Justified?*, 14 Continuing Education in Anaesthesia, Critical Care and Pain 52–56.

37 Darren Conlon et al., *Disclosure of Confidential Information by Mental Health Nurses, of Patients They Assess to be a Risk of Harm to Self or Others: An Integrative Review*, Int. J. Ment. Health Nurs. 31402539. https://pubmed.ncbi.nlm.nih.gov/31402539 (last visited Apr. 29, 2021).

38 Jack Balkin, *Information Fiduciaries and the First Amendment*, 49 UC Davis Law Review 1205–120 (2016).

39 Mason Marks, *Suicide Prediction is Revolutionary. It Badly Needs Oversight. Should we Trust Facebook to Dispatch Police to the Homes of Distraught Users?* (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) (last visited Jan. 18, 2021); Mason Marks, *Facebook Should*

Mark Zuckerberg and Prof. Zittrain from Harvard Law School on whether Facebook should be considered an 'information fiduciary' when it comes to the privacy of its clients, Zuckerberg pointed out that this resonates with the services they provide. 'The idea of us having a fiduciary relationship with the people who use our services is intuitive, . . . [Facebook's] own self-image of ourselves and what we're doing is that we're acting as fiduciaries and trying to build services for people . . . Where this gets interesting is who gets to decide in the legal sense, or in the policy sense, of what's in people's best interest' said Zuckerberg.[40]

On a high level, Facebook is also important to society. People choose to use Facebook because there is evidently some value in it. It is important to note that socially we are drifting away from the notion that online platforms such as Facebook are a neutral space and acknowledging a very active role in shaping public opinion. Therefore, in order to implement a model like the suicide risk detection system, the question of who determines what is in people's best interests is crucial.

The fiduciary framework seems to resonate with the emergent roles and obligations of social media platforms to accept more responsibility for the content being shared. If a person expresses intent to commit a violent act against themselves or against another on social media—and that intent is flagged (by either an individual or a machine) but not reported, what responsibility is borne by the flagging party?

The monitoring and filtering of social media and suicide prevention programs have also sparked debate and numerous legal problems, including important issues of freedom of speech and civil liberties. For example, images and videos of self-harming behavior are commonly shared within online communities.[41] Health professionals argue that such posts may trigger self-harming behavior in predisposed individuals and that such material should be filtered out of online communications.[42] Proponents of freedom of speech argue that not all censorship may contribute to the social isolation and stigma that people who self-harm face, potentially exacerbating the problem.[43] Marks argues that such individuals may also be deprived not only of opportunities to express themselves, but also that their Fourth Amendment rights may be violated in the context of 'warrantless searches based on opaque suicide predictions'.[44] These are essential debates, since the varying reasons and circumstances for suicides may trigger

'First do no Harm' When Collecting Health Data (Apr. 20, 2018), https://blog.petrieflom.law.harvard.edu/2018/04/20/facebook-should-first-do-no-harm-when-collecting-health-data/ (last visited Jan. 21, 2021).

40 Martha Stewart, At Harvard Law, Zittrain and Zuckerberg Discuss Encryption, *'Information Fiduciaries' and Targeted Advertisements*, CEO Visits with Students from the University's Techtopia Program and Zittrain's Internet and Society Course (Feb. 20, 2019), https://today.law.harvard.edu/at-harvard-law-zittrain-and-zuckerberg-discuss-encryption-information-fiduciaries-and-targeted-advertisements/ (last visited Jan. 18, 2021).

41 Stephen P. Lewis et al., *Non-Suicidal Self-Injury, Youth, and the Internet: What Mental Health Professionals Need to Know*, 6 Child and Adolescent Psychiatry and Mental Health 13. doi: 10.1186/1753-2000-6-13 (2012).

42 *Id.*

43 Hanna Kozlowska, *Self-Harm Images on Instagram Show How Difficult it is to Police Content Online* (Feb. 7, 2019), https://qz.com/1543307/instagram-is-introducing-sensitivity-screens-for-self-harm-content/ (last visited Jan. 28, 2021).

44 Mason Marks, *Artificial Intelligence Based Suicide Prediction* (Jan. 30, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3324874 (last visited Feb. 29, 2020).

various legal, social, and institutional implications.[45] The matter is complicated by the fact that legal frameworks for civil liberties and, the degree of Internet censorship varies from country to country. Although most democratic countries have moderate Internet censorship, other countries are able to have more control over Internet content.[46] The bone of contention is whether the public sector or the private sector should be responsible for limiting content on the Internet and how much restriction should be allowed. Social media platforms such as Facebook are generally less regulated than other forms of media such as newspapers, radio, or television. The generation and transmission of information over social networks is thus decentralized and more dynamic. Therefore, imposing restrictions on social media may clash with the fundamental rights of freedom of speech and expression such as those enshrined in the US Constitution.[47]

III. ETHICAL CONSIDERATIONS

Healthcare professionals and medical law scholars have raised concerns about entities like Facebook, which are not health care providers, seemingly providing health advice and intervention without being held to the same ethical standards as legitimate health care providers.[48] Childress and Beauchamp developed the four principles of health care ethics (autonomy, beneficence, non-maleficence, and justice) to guide the ethical provision of health and medical treatment.[49] It is expected that all healthcare providers adhere to these ethical guidelines when providing health care to individuals. Social media platforms like Facebook, though acting like a health care provider by screening, analyzing, and acting upon personal health information collected by its suicide detection algorithm, are not officially considered health care providers. Therefore, they are not obligated to follow the same set of ethical and legal guidelines as true health care providers. Facebook's program operates within a legal grey area with nearly no oversight.[50]

In one recent example of Facebook's suicide detection algorithm identifying a suicide risk, an individual was escorted to an inpatient psychiatric hospital by law enforcement for a mental health evaluation despite no previous history of mental illness or suicidal behavior and the individual's assertion that they were not experiencing suicidal thoughts.[51] Per Facebook's protocol for intervention once a suicide risk is

45 Kevin M. Simmons, *Suicide and Death with Dignity*, 5 Journal of Law and the Biosciences 436–439 (August 2018), https://doi.org/10.1093/jlb/lsy008.

46 Brian Mishara and David Weisstub, *Ethical, legal and practical issues in the control and regulation of suicide promotion and assistance over the Internet*, 37 Suicide Life Threat Behav 58–65 (2007).

47 Brian Mishara and David Weisstub, *The Legal Status of Suicide: A Global Review*, 44 International journal of law and psychiatry 54–74 (2016). The First Amendment does not apply to social media companies. However, some argue that it should or that censorship on social media violates the spirit of the First Amendment if not the law itself.

48 Mason Marks, *Artificial Intelligence Based Suicide Prediction* (Jan. 30, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3324874 (last visited Feb. 29, 2020).

49 James Childress and Tom Beauchamp, Principles of Biomedical Ethics (Oxford University Press New York 2001).

50 Mason Marks, *Artificial Intelligence Based Suicide Prediction* (Jan. 30, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3324874 (last visited 29 February 2020).

51 Natasha Singer, *Screening for Suicide Risk, Facebook Takes On Tricky Public Health Role*, N.Y. Times (June, 2019). https://www.nytimes.com/2018/12/31/technology/facebook-suicide-screening-algorithm.html (last visited Feb. 29, 2021).

detected, law enforcement was obliged to follow through with taking the individual to the hospital, precluding the individual's right to choose their own treatment course. In the case of individuals who are brought in potentially against their will for a psychological evaluation in the context of a false positive, Marks suggests that the health care ethics principle of autonomy is being violated.[52] According to Facebook's Global Head of Safety, Antigone Davis, in 2018 alone there were 3500 reports, which prompted Facebook to activate emergency responders on average of 10 times per day to carry out a wellness check.[53] Facebook staffers are even tasked with discerning whether a situation requires police intervention[54]—a prospect which, at least in some communities in the United States, is layered in the historical mistrust of law enforcement and horrific police brutality.[55] The idea of police carrying out mental health wellness checks in such communities based on a Facebook post may only serve to further exacerbate existing inequalities, and based on historical precedent, potentially even lead to violent encounters.[56]

By collecting and analyzing data on a novel public health intervention, it can be argued that Facebook's suicide detection algorithm is also a large-scale medical research project. The medical research community has strict standards for protecting the rights of research participants as outlined in the Belmont Report.[57]

IV. WIDER CONSIDERATIONS

Mental illnesses, the individuals who experience them and their families, have a long history of being stigmatized in most societies.[58] Stigma continues to be a reality for individuals with mental illness who either choose to disclose their illness or perhaps have it revealed without their consent. Because of the pervasiveness of stigma within many cultural contexts around the world, anti-stigma campaigns have not been widespread. It is within this context of mental illness stigma that Facebook's suicide prevention algorithm and accompanying intervention pose a serious not only ethical but also cultural dilemma. Depending on whom the results of a positive suicide risk detected by the algorithm are reported to, an individual with mental illness could be ostracized from their community and support systems. For example, if a family member is made privy to the information, they may choose to dissociate with the identified family member, perhaps even repudiate them. Being disconnected from family is a devastating, stressful situation for any person, but for individuals with mental illness, such isolation

52 Jack Balkin, *Information Fiduciaries and the First Amendment*, 49 UC Davis Law Review 1205–120 (2016).

53 Martin Kaste, *Facebook Increasingly Reliant on A.I. to Predict Suicide Risk*, NPR (2018), https://www.npr.org/2018/11/17/668408122/facebook-increasingly-reliant-on-a-i-to-predict-suicide-risk?t=1610820256972 (last visited Jan. 18, 2021).

54 *Id.*

55 Sirry Alang et al., *Police Brutality and Black Health: Setting the Agenda for Public Health Scholars*, 107 American Journal of Public Health 662 (2017), doi: 10.2105/AJPH.2017.303691

56 Matthew Desmond et al., *Police Violence and Citizen Crime Reporting in the Black Community*, 81 American Sociology Review 857–876 (2016), doi: 10.1177/0003122416663494.

57 Jennifer Sims, *A Brief Review of the Belmont Report*, 29 Dimensions of Critical Care Nursing 173–174 (2010).

58 Patrick Corrigan and Abigail Wassel, *Understanding and Influencing the Stigma of Mental Illness*, 46 J. Psychosoc. Nurs. Ment. Health Serv. 42–48 (Jan. 2008) and Tahirah Abdullah and Tamara L. Brown, *Mental Illness Stigma and Ethnocultural Beliefs, Values and Norms: An Integrative Review*, 31 Clinical Psychology Review 934–948 (Aug. 4, 2011)

and social upheaval may exacerbate symptoms. Especially for individuals experiencing suicidal thoughts, discord in interpersonal relationships can serve to intensify suicidal thoughts and often precipitate suicide attempts.[59] To further compound the cultural ramifications of suicide risk detection with the suicide prevention algorithm, it is not really known how well the algorithm performs across various cultural contexts and languages. Though the algorithm is trained to detect key words and phrases related to suicide risk in English, Spanish, Portuguese, and Arabic, experts in the medical field question whether the algorithm works equally when applied to different ethnic groups, genders, and nationalities.[60] Such unknowns leave room for the possibility of false positives, which can lead to numerous undesirable outcomes, included stigmatization, unnecessary psychiatric inpatient hospitalizations, trauma from stressful and potentially violent encounters with law enforcement, and exacerbation of mental illness symptoms and suicide risk.[61]

Last but not least, it is important that the focus on social media does not draw too much attention from other forms of prediction mechanisms and interventions, and in particular those that involve human responses. For instance, the impressive reduction of suicide rates in Denmark,[62] demonstrate the significance of establishing suicide prevention clinics and psychiatric emergency outreach teams that offer counseling, therapy, or visits and practical support to persons with suicidal ideation or behavior.[63] Any social media and algorithm-based strategies should carefully consider—and be linked to—human 'hands on' involvement. In that way social media-based strategies could help to further reduce suicide through supporting targeted interventions for selected risk groups.

V. CONCLUSIONS

Social media platforms, such as Facebook, have an enormous potential to greatly impact public health, for better or for worse. Suicidal behavior is influenced by social networks, and social media has expanded everyone's social network on an exponential scale. As such, social media simultaneously has the ability to perpetuate suicidal behavior while also save lives. In acknowledgment of this influential role, Facebook developed a suicide risk detection algorithm. Though an appropriate response to such a serious public health problem as suicide, the use of the algorithm raises important legal, ethical, and cultural considerations that have not yet been adequately addressed. More

59 Lindsay Sheehan et al., *Benefits and Risks of Suicide Disclosure*, 223 Soc. Sci. Med. 16–23 (Feb. 2019) and Mia Rajalin et al., *Family History of Suicide and Interpersonal Functioning in Suicide Attempters*, 247 Psychiatry Res. 310–314 (Jan. 2017).

60 Megan Thielking, *Experts Raise Questions About Facebook's Suicide Prevention Tools*—STAT (Feb. 2019), https://www.statnews.com/2019/02/11/facebook-suicide-prevention-tools-ethics-privacy/ (last visited Mar. 2, 2020).

61 Mason Marks, *Artificial Intelligence Based Suicide Prediction* (Jan. 2019), (last visited Feb. 29, 2020).

62 Merete Nordentoft and Annette Erlangsen, *Suicide—Turning the Tide*, 365 Science 725, doi: 10.1126/science.aaz1568 (describing how Danish Suicide Prevention Clinics that have 'offered counseling, therapy, and practical support to persons with suicidal ideation or behavior nationwide since 2007', has been 'linked to long-term reductions in fatal (29%) and nonfatal (18%) suicidal acts)'.

63 *Id.* (arguing that further reductions in suicide could be achieved through anonymous counseling through hotlines and targeted interventions for selected risk groups).

discourse as well as appropriate research needs to occur in order to fully understand the ramifications of suicide detection algorithms implemented by social media platforms.

Facebook may consider it wise counsel to take a page from the field of human subjects research with respect to applying standardized ethical practices. Prior to collecting sensitive data from its users, an ethical review process must be carried out. Other scholars have suggested that to neglect such ethical practice raises serious safety concerns.[64] Subsequent to this, data collection should only be done with the explicit approval of its users in the form of informed consent. The nature of informed consent in the medical field is an ongoing process. It is a continuous communication process to ensure respect and safeguard a patient's autonomy.[65] This ethical review process can be facilitated with social media platforms taking the role of fiduciaries and with the assistance of a panel of external impartial experts from various fields and diverse backgrounds. This process could turn into hard law and would at the very minimum serve to add a layer of transparency to what currently exists as an opaque process.

ACKNOWLEDGMENTS

The research for this paper was supported by a Novo Nordisk Foundation grant for a scientifically independent Collaborative Research Program in Biomedical Innovation Law (grant agreement number NNF17SA0027784) and by the Alexander von Humboldt-Stiftung, Bonn, Germany.

64 Mason Marks, *Artificial Intelligence Based Suicide Prediction* (Jan. 30, 2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3324874 (last visited Feb. 29, 2020).

65 Marcelo Corrales and George Kousiouris, *Nudging Cloud Providers: Improving Cloud Architectures Through Intermediary Services*, in New Technology, Big Data and the Future of Law 157 (Marcelo Corrales, Mark Fenwick and Nikolaus Forgó, eds., Springer 2017).