*Penultimate draft - to appear in the *The Oxford Handbook of Philosophy of Suicide,* Stellino, P. & Cholbi, M. (Eds). Please cite published version*

Suicide, Social Media, and Artificial Intelligence

Abstract: Suicide is a complex act whose meanings, while sometimes tragic, vary widely. This chapter surveys the ethical landscape surrounding algorithmic methods of suicide prevention especially as it pertains to social media activity and to the moderation of online suicide communities. We begin with a typology of suicide, distinguishing between varied goals in which suicide may factor as a means. Suicides should be understood as an act with varied eliciting desires, meanings, consequences, and ethics. Further, while many suicides may be grounded on irrational desire, we argue that a desire to suicide may nonetheless be rational in some cases. We then offer a survey of the attempts by social media corporations to create suicide prediction algorithms and launch interventions aimed at preventing suicide as well as an overview of the current ethical debate. We argue that the values embedded within algorithmic suicide prevention nudges are "thick" in a substantive sense and deserving of scrutiny. Given that debates about the rationality, meaning, and permissibility of suicide are common and subject to intracultural and cross-cultural differences, we maintain that the task of suicide prediction is ill-suited to algorithmic approaches. Despite an algorithm's ability to quantify a complex phenomenon like suicide risk and its appearance of objectivity, any suicide prediction algorithm is thoroughly subjective and involves moral judgments all the way down. Any such algorithm would necessarily impose a normative stance on suicide onto its user base that is subject to reasonable intra and cross-cultural disagreement.

Keywords: suicide prediction, suicide prevention, artificial intelligence, social media, mental illness, rationality

Susan Kennedy
Santa Clara University

Erick Jose Ramirez
Santa Clara University

In 2020 John Steen logged on to his Facebook account much as he did on many other days. He noticed that his good friend, Ronnie McNutt, was in the middle of a livestream and decided to drop in on him. None of this was unusual. Ronnie and John were friends and Ronnie wasn't new to livestreaming on the platform. Ronnie would often livestream to talk about his life with his friends and with whomever happened to be watching his stream. What happened next forever altered both of their lives. John noticed that Ronnie was acting intoxicated and that he was holding a rifle. John repeatedly flagged the video to alert Facebook moderators about its content but without success. Ronnie botched an attempt to load his rifle causing John to call local authorities in the hopes that they would be able to get to his friend in time, but they were all too late. Ronnie ended his life in the middle of the livestream, a recording of which quickly went viral (Wakefield 2020). This was not the first time that Facebook, and its parent company *Meta*, has dealt with legal and moral issues relating to its users' behaviors. Ronnie's final Facebook post, found only after his death, was a call to action: "Someone in your life needs to hear that they matter. That they are loved. That they have a future. Be the one to tell them" (Warnock 2020). Ronnie's death on the social media platform helped to draw more attention to the fact that

social media platforms like Facebook, which have many millions of active daily users, can play an important role in addressing suicides on its platform.

In this chapter we focus on a particular sort of response that social media companies have taken to address suicide and suicidal ideaion among their user bases: algorithmic suicide prevention. Specifically, we introduce a set of ethical concerns that, we believe, all algorithmic approaches must confront. We begin by investigating the nature and meaning of suicide. Suicides are complex events about which there are significant intra and cross-cultural interpretations. We then focus on "The Orthodox View" of suicide among Western bioethicists and mental health practitioners. The Orthodox View of suicide is that suicidal desires are grounded in mental illness and that, because of this, suicidal ideation (and suicide itself) requires preventative psychiatric attention. We challenge this view on several fronts. We claim that suicide can, at least on some occasions, be rational. That suicidal desire is not necessarily indicative of an underlying mental illness and that the mere presence of a diagnosable mental illness need not undermine the rational competency of someone with suicidal desires. We do this in order to argue that algorithmic approaches to suicide introduce several ethical problems ranging from concerns over data collection and privacy to the imposition of culturally-specific thick values on the nature and appropriateness of death. While suicide is often a tragic consequence of mental illness, and prevention of such suicides a laudable and important goal, algorithmic approaches to suicide prevention are largely unethical. We close the chapter with suggestions for ethical alternatives to the use of algorithms.

A Typology for Suicide

Among mental health practitioners, the Orthodox View of suicide is that suicide is the result of irrational desires and beliefs which have their origin in underlying mental illness (Hoven, Mandell, & Bertolote 2010; Swanson, Bonnie, & Appelbaum 2015; Nelson and Ramirez 2017). Suicide is largely framed as a public health problem about which social media companies are in an especially advantageous position to help solve (Luxton, June, & Fairall 2012). The Orthodox view grounds this way of thinking about suicide by tying together suicide with illness (that suicide results from mental illness) and tying mental illness with irrationality. Conflating these concepts is then understood to give license to mental health practitioners, and social media companies, to intervene on behalf of those who express (or act on) desires to suicide.

While it is sometimes true that individuals who suicide are mentally ill, irrational, and that in desiring suicide act counter to their own interests in this section we argue for several claims. First, we argue that a desire to end one's life is not necessarily a symptom of underlying mental illness and thus that such a desire is not intrinsically irrational. Second, we claim mentally ill persons can, in some instances, rationally desire to end their lives (that is, that mental illness does not always preclude rational agency). It is possible, we claim, for suicide to be be in line

with a person's autonomously expressed desires and not always counter to their own interests. In these cases, it can be wrong to prevent such a person from expressing or acting on these desires.

We support these claims by offering a typology of the multiple meanings that suicide may have in the context of an individual's life. Suicide, on this view, is not a purely descriptive term to describe an action but is, instead, deeply intertwined with normative concepts of rationality and ethics. We then discuss the Orthodox View before arguing its adherents wrongly tie together suicidal desire with mental illness and irrationality.

Meaning of Suicide

We distinguish between three motivations that can ground desires to suicide: radical autonomy, one's own well-being, and the well-being of others. Table 1 below summarizes these motivations. We understand remedial suicides to include a wide array of situations in which individuals judge that their continued existence would be against their interests. Terminally ill patients who withdraw treatment to hasten their death (Brody 2007) would fall under this category as would instances in which people prefer death to captivity or torture.

Other-directed suicides, as the name implies, use suicide to convey political and/or moral messages to others. Other-directed suicides range widely to include the likes of Kendrick Castillo, an 18 year old high school student, who was killed preventing a gunman from murdering his fellow classmates in 2019 (Brito 2019). Similarly, Buddhist monk Thích Quảng Đức's death by self immolation in 1963 in an act protest against his government's repression of Buddhist practice is an instance of an other-directed suicide (Sanburn 2011). A patient needing expensive long-term care who chooses to forego treatments in order to spare their family the cost of their medical care also engages in other-directed suicide as does a spouse who ends their life in order to hurt their partner. The moral value of such suicides will depend on the background normative framework one applies to such cases (Young 2002).

Table 1. A typology for suicidal motivation

	Per-Se Suicide	**Remedial Suicide**	**Other-directed Suicide**
Cause	Expression of individual autonomy, responsibility, or ultimate values	A judgment that one's well-being is improved by ending one's life rather than continued living	A judgment that one's death is instrumental to improving, or harming, another's situation
Example	Performance artist ending their life on stage as an existential	A terminally ill patient requesting, and taking, a lethal	A soldier who intentionally jumps onto a grenade in order

	statement	dose of medication	to save their comrades

Lastly, although what we've described as "per-se" suicides are rare, they can occasionally happen. In 2011 19 year-old Kipp Rusty Walker ended his life on stage after performing a song called "Sorry For All the Mess" in what many have interpreted as performance art (James 2011). Existential philosopher Albert Camus famously argued that "[t]here is only one really serious philosophical question, and that is suicide" (Camus 1955). The decision to suicide or continue living, in his view, is always present, to everyone. In the face of existential absurdity, suicide can, at least sometimes, be understood as an affirmation of radical freedom. In saying this we do not deny that many instances of suicide can be irrational and counter to someone's interests. Suicidal motivations are never purely descriptive. Instead they involve normatively complex interactions between a person's moral, spiritual, and emotional values. In offering this typology we mean to show that the motivations behind suicide can be complex. This complexity also affects our understanding of the rationality (or irrationality) of suicide. Any assessment of the ethics of suicide prevention only makes sense in the context of this complexity.

The Orthodox View of Suicide

Though an individual's motivations for suicide may be per-se, other-directed, or remedial, an understanding of a person's motivations for suicide will not be enough to assess the rationality or irrationality of the desire. On the Orthodox view *all* desires for suicide are irrational. To see how such a view might arise, it can be helpful to explore how suicidal ideation is pathologized in the two dominant global nosology manuals: the *Diagnostic and Statistical Manual of Mental Disorders* (the DSM) and the *International Classification of Disesases* (American Psychiatric Association 2022; World Health Organization 2022). With respect to suicidal desires and ideation both manuals cover similar ground and diagnostically they are intended to be interchangeable (i.e., diagnostic reference numbers in the DSM commonly include their ICD equivalent). For the sake of brevity, we'll thus focus on the DSM.

Although thoughts of suicide are associated with many discrete illness categories by the DSM, two diagnostic categories in particular are helpful to show how the Orthodox View is embedded in the document. Major Depressive Disorder (MDD) is often commonaly associated with suicidal ideation and an increased risk of attempts to suicide (Dong et al. 2019). MDD is characterized in the DSM as having, among its essential features:

> a period of at least 2 weeks during which there is either depressed mood or the loss of interest or pleasure in nearly all activities…The individual must also experience at least four additional symptoms drawn from a list that includes changes in appetite or weight, sleep, and psychomotor activity; decreased energy; feelings of worthlessness or guilt; difficulty thinking, concentrating, or making decisions; or *recurrent thoughts of death or*

> *suicidal ideation or suicide plans or attempts*…The episode must be accompanied by clinically significant distress or impairment in social, occupational, or other important areas of functioning. (APA 2022)

Clinicians are not required to assess the rationality of an individual's feelings of worthlessness or guilt or their suicidal ideation. The mere presence of such thoughts is understood to be symptomatic of an underlying dysfunction. When discussing suicidal ideation associated with MDD, the document focuses on remedial and other-directed causes:

> Motivations for suicide may include a desire to give up in the face of perceived insurmountable obstacles, an intense wish to end what is perceived as an unending and excruciatingly painful emotional state, an inability to foresee any enjoyment in life, or the wish to not be a burden to others. The resolution of such thinking may be a more meaningful measure of diminished suicide risk than denial of further plans for suicide. (APA 2022)

Critics of the APA's removal of the "Bereavement Exclusion" from the diagnosis of MDD, for example, have argued against pathologizing all instances of intense grief and suicidal ideation (Horwitz 2015). Although the Orthodox View is integrated deeply into nosologies like the DSM and ICD, and thus built into algorithmic suicide prevention tools, the view itself is not defended. We now subject the Orthodox View to critique.

Rational Suicide

An assumption built into the Orthodox View is that desires for suicide are grounded in illness and that, on that basis, compromise a person's agency and autonomy enough to allow clinicians to doubt the competence of the person giving voice to the desire. The concept of rationality (Broome 2007; Langsam 2008; Weatherson 2019), and related concepts like medical competence (Culver and Gert 2006), have a long history. Though accounts differ, we argue here that rationality is best understood in terms of the intersubjectively valid exchange of reasons (Fischer and Ravizza 1998; Broome and Bortolotti 2009; Kiesewetter 2017; Nelson and Ramirez 2017).

To say that someone is rational, on such a theory, is to say that they are able to articulate reasons for their beliefs that a community of other agents would find intelligible as grounds for the belief. Note that intelligibility does not mean agreement. A rational agent can articulate reasons for their beliefs that others might understand but nonetheless reject. Note further that communities may be geographically or linguistically isolated from others.

For example, a childfree adult may understand a friend's desire to have a child in order to leave a legacy that continues after their death while nonetheless believing that it's better for them to

remain childfree. An atheist can find intelligible a theist's rationale for martyrdom while doubting the metaphysical assumptions upon which that claim is grounded. Similarly, the mere offering of reasons does not, by itself, show that an agent satisfies the criteria for being intersubjectively rational. Matthew Broome and Lisa Bortolotti (2009) offer an example of delusional beliefs

> in which the subject with delusions comes up with understandable reasons in support of the reported delusional belief, but such reasons are not intersubjectively good reasons...A good example is that of a woman who claims that her blood is being injected out of her body in her sleep because she has spots on her arms. When the interviewer says that the spots are freckles and that he has them too, she agrees that the spots are similar to freckles, but continues to believe that she is being injected. (35)

In a case like this, it's difficult to imagine, without being offered more information, why the presence of freckles on someone's arm would count as a reason to believe that blood is being taken from them. The bare offering of reason is not sufficient for intersubjective understanding.

Motivations for suicide are many and varied. In some instances, such motivations may fail to be intersubjectively valid because they don't portray realistic information about the world. Someone who desires suicide based on a misunderstanding of their prognosis (i.e., they believe they have a fatal disease that is actually easily treatable) or their life circumstances (i.e., they believe that nobody cares about them but actually have loved ones who care deeply for them) may be acting on the basis of such a mistake. Additionally, a desire for suicide may be grounded in rationality-undermining mental illness in the way that adherents of the Orthodox View assume and in much the same way as the delusion example above.

Despite these possibilities, it remains equally probable that desires for suicide can be intersubjectively valid and thus rational. Similarly, some suicidal desires occur *in the presence of* but not *causally connected with* mental illness. Laws permitting medicalized suicide (actively or passively) vary widely across the Western nations (Downie 2016; Ebbott 2010) but the existence of the practice indicates that, in at least some cases, a desire for suicide is seen as an intersubjectively valid response to one's life circumstances (even if a person who makes such a desire satisfies the diagnostic criteria for one or another mental illness). An individual who, after years of trying, has exhausted their treatment options may, as a result, evince symptoms of major depressive disorder. A desire for suicide, in such a circumstance, may nonetheless be rational. Additionally, it's important to realize that attitudes about the permissibility of suicide can vary widely within a single Amerian state (Periyakoil, Kraemer, & Neri 2016) and between states (Range et al. 1999). Globally, attitudes about the permissibility of suicide can range from outright illegality (Adinkrah 2016) to social and cultural acceptance (Kim & Park 2014; Young

2002). One study, comparing Korean student's attitudes with their American counterparts, found a clear *normative* disagreement about suicide prevention:

> The findings that students in Korea tend to agree that the duration of suicidal process is long, suicide is predictable, and people communicate their suicidal intent to others is somewhat surprising because these indicate a belief in the possibility of suicide prevention. However, the students also reported that they believe people do not have a right to prevent suicide, which is an *ethical* issue. It may be that although the students believe it is possible to prevent suicide, they do not believe it is ethical, and this attitude may in turn hinder collective suicide prevention efforts. (Kim and Park 2014)

Such disagreements are likely to introduce further ethical issues with suicide prevention interventions (we discuss these in more detail in a later section). In order to assess whether a suicide-referencing post on a social media platform requires intervention therefore requires a complex assessment. The following criteria are especially relevant (Nelson and Ramirez 2017):

1. Whether a user's post's is rationally intelligible
2. Whether the post or user is able to convey realistic information and judgment about their life and world
3. Whether the user is in a lucid state of mind, expressing intelligible emotion, and authentic will
4. Whether the user's post is congruent with their values and critical interests

Fundamental to our view is that assessments about the rationality of suicidal beliefs or desires, like assessments of rationality itself, are normative (Way 2010; Kiesewetter 2017). Any attempt to adjudicate whether a suicidal desire expressed on a social media platform should receive an in intervention will thereby include assumptions about its irrationality and moral permissibility. Because such posts are deeply intertwined with a user's values, identity, and culture, the rationality of these posts is not easily subject to algorithmic assessment. Without building in intracultural and cross-cultural variability into algorithmic suicide prevention features, social media platforms that use such features will impose a single, largely Western and American, conception of suicidal desire onto all of its users.

Social Media and AI-driven Suicide Prevention Tools

The identification and intervention of suicide risk has traditionally been the purview of healthcare professionals. However, in recent years, social media platforms have emerged as a major contributor to suicide prevention efforts. In both contexts, there has been increasing interest in the use of AI-driven methods to predict suicide risk (Bernert et al. 2020). For example, AI tools that leverage patient data from electronic medical records are currently being explored

as a diagnostic support tool for healthcare providers to achieve more accurate assessments of suicide risk (Walsh, Ribeiro, & Franklin 2017; Lejeune et al. 2022). In the context of social media, companies are producing AI tools to monitor users' activity to predict suicide risk in order to target various prevention measures accordingly.

Given the rising suicide rates in recent years, particularly among adolescents, and the assumptions embedded within the Orthodox View that suicides should always be prevented, suicide is increasingly being recognized as a pressing public health issue requiring attention and solutions (David-Ferdon et al. 2016; World Health Organization 2016). The negative influence that social media may have on suicide-related behavior has been the subject of numerous studies (Luxton, June, & Fairall 2012; Memon et al. 2018). Nonetheless, the advent of AI-suicide prediction on social media platforms has been generally well-received and, in some cases, explicitly sought after. For instance, Facebook employees reported that "a number of suicide prevention institutions came to us and highlighted our unique position to help tackle this problem" (Gomes de Andrade et al. 2018, 678). Research has shown that social connectedness is one of the key protective factors against suicide (Centers for Disease Control and Prevention 2008), and social media platforms possess the infrastructure (e.g., behavioral, technological, and social) needed to launch effective interventions in this regard. For example, social media companies can connect an individual in distress with people who care about them, direct them towards supportive online communities, or present them with other resources and services that can provide help (Luxton, June, & Fairall 2012; Rice at al. 2016).

Additionally, direct involvement in suicide prevention efforts by social media platforms has been argued to be beneficial as these interventions have the potential to reach at-risk users who might not engage with licensed healthcare providers (D'Hotman and Loh 2020; D'Hotman, Loh, & Savulescu 2021; Halsband and Heinrichs 2022; Rice at al. 2016). The increasing use of social media for communication, particularly among adolescents, means that an individual's expression of distress and suicidality may be present on social media before, or as an alternative to, disclosure to medical professionals (Belfort, Mezzacappa, & Ginnis 2012; Marchant et al. 2017; Pourmand et al. 2019). Thus, social media companies are arguably in a favorable position to intervene before medical professionals have the chance. Moreover, according to the Orthodox View, users at-risk for suicide are assumed to have mental illness and, being irrational, are unable to identify their need for help. As a result, many at-risk individuals fall outside the reach of medical professionals' interventions. On social media platforms, however, these at-risk individuals can be met with suicide prevention efforts they would not otherwise encounter and which, on the Orthodox view, are paternalistically justified.

Social media companies have access to large datasets from their widely-popular digital platforms, which can be used to train and develop algorithms to predict suicide risk. The use of AI to assessing suicide risk is significant as it enables these companies to *automate* a suicide prevention tool, making it more efficient to implement an intervention on a global scale. Unlike

the interpersonal relationship between patients and healthcare providers which are grounded in a particular context, community, and culture, social media platforms exhibit a "one-to-many" relationship with their users. An algorithm produced by a social media company can be implemented across millions of diverse users worldwide, thereby imposing a single, seemingly objective, normative conception of what constitutes suicidal risk as well as which individuals are an appropriate target for intervention.

As one might expect, the highly concentrated influence of social media companies has prompted discussion about the moral permissibility of AI suicide prevention tools. In the remainder of this section, we offer a survey of the attempts made by social media companies to algorithmically predict and prevent suicide as well as an overview of the current ethical debate.

Ethics of Suicide Interventions

Once an algorithm identifies an individual as both at-risk and as a valid target for intervention, social media companies are believed to have prima facie justification for launching a suicide prevention options based on Orthodox View which ties suicide to mental illness, deems it irrational, morally wrong and therefore warranting prevention (Halsband and Heinrichs 2022). The goal of interventions aimed at individual users may be to provide resources to help them overcome their suicidal desires or, alternatively, to directly impede a suicide attempt. The former are referred to as "soft touch" interventions, while the latter are referred to as "hard touch" interventions. There is also a third category of interventions that operate at the population level and are not targeted at individuals. Given that the ethical concerns associated with each of these intervention categories are distinct, we will address each in turn.

a. Individual Interventions - Soft Touch

The concept of a “nudge”, popularized by Richard Thaler and Cass Sunstein (2009), maintains that small, non-intrusive changes in an individual's environment can lead to changes in behavior. Nudges are built into social spaces of all sorts and include decisions about how to display products in a grocery store, the presence of speed bumps on a road surface, and the inclusion of caloric information on menus. The ethics of implementing nudges into a choice-architecture (a space in which choices are made) stresses the importance of value-alignment and transparency.

In the context of suicide prevention, soft touch interventions are a type of nudge that aims to gently encourage individuals to seek help without restricting their freedom of choice. For example, when users search for topics identified with suicide risk, Snapchat provides users with resources via a “Here for You” page (Snapchat 2022) and Twitter’s #ThereIsHelp feature recommends suicide prevention resources by pinning them to the top of users’ search results (Meche 2020). On Facebook, users whose posts have been reported for risk of suicide or

self-harm are presented with various resources including support hotlines, online chats, and additional information (Gomes de Andrade et al. 2018). In all of these cases, the use of such resources remains purely optional.

Users may not have the option to opt-out of their social media posts being screened for suicide risk, as is the case with Facebook's suicide prevention tool, but soft-touch interventions still allow for freedom of choice on behalf of the user. Like nudges, these interventions may be considered a morally permissible way to empower individuals to seek social support and professional assistance without infringing on their autonomy. Unlike involuntary commitment or forced treatment in the medical context which can be seen as a violation of an individual's autonomy, soft touch interventions are less likely to be perceived as threatening or coercive and may be more effective in encouraging individuals to seek help.

b. Individual Interventions - Hard Touch

In contrast to soft touch interventions, hard touch interventions are more intrusive and coercive. In these cases, social media companies will arrange for individuals who are identified as at-risk to receive some form of real-world intervention. This could include contacting local services such as first responders to check on the individual or arranging for involuntary commitment or forced treatment. In these cases, the intervention is not optional and individuals do not have any say in whether or not it takes place. One of the most prominent examples of a hard touch intervention is Facebook's "Wellness Checks." If a person is determined to pose an imminent danger to themselves based on their social media posts, Facebook uses geolocation data to contact local services such as first responders who can attempt a real world intervention (Card 2021). Based on a report from Facebook's Global Head of Safety, Antigone Davis, 100 wellness checks had been conducted in the first month of the initiative's implementation and this number rose to 3,500 throughout the rest of the year (Kaste 2018).

According to the Orthodox View, individuals at risk of suicide are assumed to have mental illness such that their suicide-related desires and behaviors are non-autonomous and not deserving of respect. As such, even coercive hard-touch interventions may be considered a morally justified form of paternalism (Halsband and Heinrichs 2022). Nevertheless, ethical concerns remain as it has been argued that these interventions may lead to violent confrontation, involuntary hospitalization or incarceration (Marks 2019). While there is little information about the outcomes of Facebook's Wellness Checks, Marks presents several cases where a wellness check initiated by a third party on the basis of social media content resulted in police resorting to lethal force in their interaction with the person in distress. In addition, Marks considers how wellness checks that take place in countries outside of the U.S., where attempted suicide is a criminal offense, may result to the criminal prosecution and incarceration of an at-risk individual. Because the meanings, and value, of suicide vary across populations and nations, hard touch

interventions risk imposing one culture's views on suicide (or a dominant sub-culture's view in the case of multinational states) onto everyone.

A different set of hard touch interventions involve content moderation. Major social media platforms such as Facebook, Instagram, and TikTok all have some form of ban on pro-suicide content (Chan 2021; *Instagram: New Tools to Ban Self-Harm and Suicide Posts* 2020; Reuters Staff 2019). If an algorithm deems user-generated content to indicate suicide risk or otherwise express a favorable, uncritical view towards suicide, the content is marked for removal from the platform. Users cannot opt out of having their content removed or restricted, although they may attempt to bypass algorithmic censorship, a phenomenon referred to as "algospeak," by using code words such as "becoming unalive" to discuss suicide (Lorenz 2022). The ethics of moderating suicidal content closely resembles broader discussions about content moderation on social media, with concerns about censorship harms and infringing users' freedom of expression. However, one criticism of content moderation that is specific to the case of suicide is that it may turn out to be a counterproductive measure for the purpose of suicide prevention. In particular, Marks (2019) has suggested that content moderation may be effective in some regards, but that "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" (119).

c. Population-level Interventions

Population-level suicide interventions are unlike soft and hard touch interventions in that they don't aim to identify individual users for suicidal ideation or risk. Instead, these interventions aim to identify *communities* with a high risk of suicide in order to inform the allocation of mental health and suicide prevention resources by the government. For example, in 2018 the Public Health Agency of Canada announced a pilot project with Advanced Symbolics to predict regional suicide rates by using AI to monitor social media activity. According to the terms of their recent contract, Advanced Symbolics would assist the Canadian government in defining suicide-related behavior, taken to include "ideation (i.e., thoughts), behaviors (i.e., suicide attempts, self-harm, suicide) and communications (i.e., suicidal threats, plans)" (Public Health Agency of Canada 2018). Following this, an AI-tool would be developed to identify patterns of suicide-risk using anonymized data from social media.

Many of the ethical concerns associated with suicide interventions aimed at individuals do not apply to population-level initiatives. Since they do not require the identification of individuals, they are assumed to be less ethically fraught as they do not invoke issues surrounding privacy and consent (D'Hotman and Loh 2020; D'Hotman, Loh, & Savulescu 2021). Moreover, since these interventions do not aim at individuals, they are also assumed to avoid ethical concerns about autonomy, paternalism, and safety. As a result, population-level interventions are usually

thought to be an ethical means to advance public health outcomes relating to suicide. We will raise issues for all of these assumptions.

Ethics of Social Sucide Prediction

A distinct thread of the current ethical debate pertains not to the *type* of intervention aimed at individuals deemed at risk, but rather to use of AI for the task of suicide prediction. Ethical concerns already raised about AI-suicide prediction are mostly indistinguishable from concerns related to the the use of AI more broadly. This may be explained by the fact that there are currently existing guidelines about the ethical use of AI which can be used to frame an ethical analysis of suicide prediction algorithms. For example, D'Hotman, Loh & Savulescu (2021) adopt a top-down ethical approach that uses the UK Government's 'Data Ethics Framework' and 'Artificial Intelligence: Australia's Ethics Framework' to identify the ethical implications of using AI for suicide prediction (103).

The use of AI for suicide prediction requires the collection and utilization of personal data from social media users, which has been met with ethical concerns about privacy, security, and informed consent (D'Hotman, Loh & Savulescu 2021; Marks 2019; Celedonia et al. 2021). While privacy interests are generally thought to be worthy of protection, infringements on privacy may be considered justifiable in the context of suicide prediction. In line with the Orthodox View of suicide, at-risk individuals are assumed to be suffering from mental illness and unable to make fully autonomous decisions. For this reason, infringing on privacy for the purpose of suicide interventions may be characterized as a morally justified form of weak paternalism. This being the case, the more salient concerns about privacy are taken to be those tied to security and consent issues which include how user data is protected, who has access to this information, and whether individuals have given their consent for their data to be used in this way. Unlike AI-suicide prediction in the U.S. medical context which is subject to review by IRBs and involves safeguards for human research subjects and data protections outlined in the Health Information Portability and Accountability Act (HIPAA), the algorithms developed by social media companies are not subject to the same laws, regulations and ethical standards (Marks 2019). To minimize opportunities for abuse, it has been argued that social media companies should be subject to greater oversight and regulation (Marks 2019; D'Hotman and Loh 2020; Celedonia et al. 2021).

Additionally, the use of AI, particularly machine learning techniques, has generated ethical concerns regarding the transparency and explainability of algorithmic decision-making, individuals' capacity to challenge these decisions, and the presence of accountability measures in case the algorithm is incorrect (D'Hotman, Loh & Savulescu 2021). Since social media companies exist in the private sector and their suicide prevention tools may be considered proprietary trade secrets, commentators have expressed concerns about the lack of information

available to evaluate the accuracy, effectiveness and overall innerworkings of these algorithms (Marks 2019; D'Hotman, Loh & Savulescu 2021). Without access to information about how the algorithms work, individuals cannot effectively challenge decisions that may result in inaccurate or discriminatory outcomes.

Algorithmic Suicide Prediction and the Normativities of Suicide

While we agree that the ethical concerns outlined so far are worthy of consideration, our approach diverges from the current debate by reversing the emphasis. That is to say, instead of solely focusing on AI as a source of concern and treating the task of suicide prediction as an incidental matter, our assessment focuses on whether the particular task of suicide prediction is inherently ill-suited to algorithmic systems. We argue that a major ethical concern that has been overlooked so far is the way controversial, normative judgments about suicide are built-in to algorithmic suicide prediction tools. More specifically, each algorithm must operate on a definition of suicide risk in order to identify which individuals should receive some intervention, and this process necessarily involves assumptions about the irrationality and moral permissibility of suicide. Unlike other ethical concerns relating to the use of AI for suicide prediction such as privacy and autonomy which can be mitigated to some extent, we take the normativity of suicide to present a fundamental ethical challenge to any AI-suicide prediction and intervention efforts.

Since algorithms offer the appearance of objectivity, there is often a failure to appreciate the subjectivity involved in the process of data mining and the development of algorithms. In particular, the creation of any suicide prediction algorithm will require decisions to be made with respect to how the target variable should be defined, how training data should be labeled, which features should be selected, and which interventions should be deployed on the basis of the algorithm's outputs. As argued by Solon Barocas and Andrew D. Selbst (2016), "many of these are normative judgements in disguise, about which there is not likely to be consensus" (676). In the context of algorithmic suicide prediction, conceptions of mental health, rationality and morality in relation to suicide are subject to intracultural and cross-cultural disagreement. As a result, the algorithms developed by social media companies are likely to impose a narrow conception relating to the normativity of suicide onto millions of diverse users worldwide. In what follows, we briefly outline some of the critical steps in the data mining and AI development process to highlight where normative judgments are likely to arise.

The first step in the process of developing an AI tool for suicide prediction requires specifying the problem to be solved. More specifically, a target variable must be defined which identifies the outcome of interest, such as a prediction about which social media users are at-risk of suicide. This will involve a decision about what constitutes suicide risk which is subject to genuine disagreements that cannot be easily resolved. For example, even if it is assumed that all suicides should be prevented there may nevertheless be disagreements about how to define suicide risk. A deeper worry, however, is that "There is no stable ground upon which to judge the

relative merits of definitions because they often reflect competing ideas about the very nature of the problem at issue" (Barocas and Selbst 2016, 715). For example, if the assumption that all suicides should be prevented is itself subject to disagreement, then settling on a definition of suicide risk will be made even more difficult if not impossible. This kind of disagreement about the very nature of the problem seems likely, especially considering that some forms of suicide, namely euthanasia and/or physician assisted suicide, are legalized in some countries and would not be an appropriate target for intervention. Nevertheless, the social media company developing the algorithm will have the final say about what constitutes suicide risk, and this may or not involve input from mental health professionals or the populations targeted by algorithmic intervention.

After a target variable has been defined, a dataset will need to be compiled in order to train an algorithm to identify suicide risk. For example, Facebook's suicide prediction algorithm was trained using a pre-existing data set containing user posts that had been reported for suicide risk to their Community Operations team (Gomes de Andrade et al. 2018). In order to improve the accuracy of the algorithm, they later decided to include positive examples of suicide risk as well as negative examples which consisted of user posts that reviewers from the Community Operations team ultimately determined not to express suicidal desires. Determining whether a user's post contains a serious expression of suicidal desire or simply a sarcastic remark, particularly in the absence of any input from the individual user about their intentions, involves subjective judgments on behalf of those developing the algorithm. Aside from the ways that the developer's views may influence the algorithm, there is also the potential for societal prejudices reflected in the training data to be reproduced by the algorithm. For instance, if Facebook's training data only included posts from users in the U.S., then the data would most likely be representative of the Western, Orthodox View and not representative of other cultural views on suicide. As a result, the algorithm may result in discriminatory outcomes for users whose views are not represented in the dataset. In short, decisions about which people and which views are included in the dataset, and which are not, will influence the algorithm's performance.

In addition to compiling a training dataset, the development of an algorithm involves a feature selection process that determines which attributes of the data will factor into the algorithm's assessment. Returning to Facebook's algorithm as an example, the features taken to be predictive suicide risk initially included select words associated with suicidal ideation but, following a consultion with experts, came to include features such as the time of day the content was posted online or other users reactions to the post (Gomes de Andrade et al. 2018). It is worth noting that, due to privacy considerations, Facebook decided against using features predictive of suicide risk that would be tied to an individual user and instead only focused on attributes of the content posted on the platform. This decision to limit feature selection to content-only may better preserve users' privacy, but it raises further questions about the algorithm's ability to assess whether a user is expressing a rational desire to suicide that does not warrant intervention. More specifically, the exclusion of features tied to the individual would make it impossible to provide

an assessment of suicide's rationality and permissibility in line with the criteria offered by Lawrence Nelson and Erick Ramirez (2017), which requires, for example, an assessment of whether the user's post is congruent with their values and critical interests. Overall, the feature selection process involves making decisions about which features, from the limited number of attributes available in the dataset, can capture a complex phenomenon such as suicide which admits of various meanings and motivations.

Following the development of a suicide prediction algorithm, a choice will need to be made about which intervention should be deployed on the basis of the algorithm's outputs. In cases where the intervention is aimed at preventing suicide, there exists a judgment, either implicit or explicit, that expressions of suicidal desires are inherently irrational and problematic. Although population-level interventions are generally taken to be less problematic than interventions aimed at individuals, either type will necessarily involve a normative judgment about suicide's permissibility. As we have argued, the view that all suicides ought to be prevented is controversial and subject to disagreement.

Values Embedded in AI

Debates about the rationality, meaning, and permissibility of suicide are common and subject to intracultural and cross-cultural differences. Consequently, the values embedded within suicide prediction algorithms are "thick" in a substantive sense and deserving of scrutiny. Insofar as social media companies commit themselves to respecting value-pluralism and multiculturalism, algorithmic interventions would starkly violate such commitments if reasonable intra or cross-cultural disagreements exist about the permissibility or rationality of suicide. As it currently stands, the algorithms adopted by social media companies would seem to have built-in values associated with the Orthodox View of suicide which are subsequently imposed on all social media users regardless of their ideological and/or cultural values with respect to suicide. Soft-touch interventions assume suicidal desires are irrational and morally wrong, content moderation of pro-suicide content discounts rational and permissible cases of suicide, and hard-touch interventions such as Facebook's Wellness Checks assume suicide is a form of mental illness, irrational, and ought to be prevented.

As Celedonia et al. (2021, 10) note, "it is not really known how well the algorithm performs across various cultural contexts and languages," but the presence of cross-cultural disagreement suggests that suicide prediction is ill-suited to algorithmic approaches. Recall that the criteria offered earlier on the nature of the rationality of suicide make it possible that suicidal desires are rational if, among other things, they are intersubjectively recognized. Recall further that even in many Western contexts that desires for suicide can be rational using these criteria (e.g., euthanasia). The problem of value imposition becomes even more severe in cross-cultural and multi-national contexts. For example, consider how Western views of suicide differ from those in Japan:

> In the West, suicide has been seen as a violation of social trust, as an act to evade social responsibility, and as an act against one's duty to God. However, in Japan, the act of suicide is perceived as acceptable, good, and rational, especially … 'when used to take social responsibility for one's acts.' The act of suicide, therefore, has a redemptive, moral quality to it. (Young 2002, 417)

A social media post expressing suicidal desires is deeply intertwined with a user's values, identity, and culture. While the expression of suicidal desires may be thought to warrant intervention in the U.S., this conclusion would not seem to follow in Japan. Similarly, many Koreans believe it is wrong to prevent someone from suicide even when a desire is made known in advance and where it *could* be preventable (Kim and Park 2014). A social media post containing suicide-relevant content may thus be both rational and deserving of respect if, on the basis of criteria offered earlier, it is intersubjectively understandable, was posted while the user was lucid, contain accurate information about the user's life and world, and is congruent with that user's individual or culturally grounded values. Whether or not such a post should be flagged by an algorithm and become the target of some intervention (and, if warranted, what that intervention should be) are questions that will require consideration of the relevant culture and geographic context to be answered appropriately. Insofar as a social media company is committed to respecting difference, especially differences grounded in moral or cultural disagreements about value, then algorithmic interventions come up short.

It might be thought that in the case of suicide prevention that "it's better to be safe than sorry" and that interventions may be justified on such a basis. Conflicts between the values of algorithmic interventions and the values of the people they target, it might be thought, could be nonetheless justified on the basis of (usually Western) ethical principles like these. However, when conflicts occur between the interests of those who are engineering a nudge and those who are being nudged, ethicists often

> insist that when conflicts of interest occur and when incentives cannot be lined up clearly, nudging is only permissible when the choice architect's design intentions are transparent and capable of being monitored. Nudges that cannot be made transparent and public thus are impermissible, as are ones that reflect racist, sexist, or other oppressive agendas. (Selinger and Whyte 2010, 467)

Algorithmic interventions, given their structure, are neither transparent or easily capable of being monitored with significant structural changes to their current design. It's also arguable that large-scale impositions of one set of geographic (Western) and culturally bound (North American) values about the nature of suicide (The Orthodox View) onto a global user base is an irreducible oppressive agenda. To implement an algorithmic suicide prevention intervention in

this context would raise significant moral and political issues and it's not clear that these interventions can be modified so as to address these issues without undermining their ability to target and intervene upon users expressing suicidal desire.

It is also worth highlighting how views on suicide can vary as much *within* cultures as it does across cultures such that a suicide prediction algorithm developed in a culturally-sensitive and context-dependent manner may still prove problematic. For example, attitudes towards the permissibility of physician assisted death (PAD) in the U.S. exhibits variability across different ethnic groups: "In California, 75.6% of non-Hispanic whites…were in support of PAD compared to 59.6% of African Americans" and 71.6% of Hispanics in California were in support of PAD compared to 63.6% of Hispanics in Hawaii (Periyakoil, Kraemer & Neri 2016, 1062). Without tailoring an algorithm's outputs to each individual, the values embedded in a suicide prediction algorithm will impose a single, normative stance on suicide that inevitably conflicts with the diverse views on suicide held by social media users.

Conclusion

Unfortunately, any suicide prediction algorithm will necessarily contain thick embedded values pertaining to suicide's connection to mental health/illness as well as its rationality and permissibility. In the process of developing a suicide prediction algorithm, moral judgments will need to be made about which posts should be considered genuine expressions of suicidal desires and which should not for the purpose of labeling training data, which views on suicide should be represented in the dataset and in what proportion, what threshold of suicide risk should trigger an intervention, which type of intervention should be deployed and what, specifically, the intervention should entail. Even the most fundamental question of whether or not a social media company should develop a suicide prediction algorithm, regardless of what form it takes, will involve a normative judgment about what *should* be the case. Despite an algorithm's ability to quantify a complex phenomenon like suicide risk and its appearance of objectivity, any suicide prediction algorithm is thoroughly subjective and involves moral judgments all the way down.

Adinkrah, Mensah. “Anti-Suicide Laws in Nine African Countries: Criminalization, Prosecution and Penalization.” *African Journal of Criminology and Justice Studies: AJCJS* 9, no. 1 (2016): 279-292

American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-5-TR*. Washington, D.C: American Psychiatric Association, 2012.

Barocas, Solon, and Andrew D. Selbst. “Big Data’s Disparate Impact.” *California Law Review* 104, no. 3 (2016): 671–732.

Belfort, Erin L., Enrico Mezzacappa, and Katherine Ginnis. “Similarities and Differences Among Adolescents Who Communicate Suicidality to Others via Electronic Versus Other Means: A Pilot Study.” *Adolescent Psychiatry* 2, no. 3 (2012): 258–62.

Bernert, Rebecca A., Amanda M. Hilberg, Ruth Melia, Jane Paik Kim, Nigam H. Shah, and Freddy Abnousi. "Artificial Intelligence and Suicide Prevention: A Systematic Review of Machine Learning Investigations." *International Journal of Environmental Research and Public Health* 17, no. 16 (2020): 5929. https://doi.org/10.3390/ijerph17165929

Brito, Christopher. “‘Hero’ Student Killed Trying to Save Classmates in Colorado School Shooting Was Just Days From Graduation.” CBS News, 8 May 2019. https://www.cbsnews.com/news/colorado-shooting-kendrick-castillo-stem-school-highlands-ranch/

Brody, Howard. “Assisted Death - A Compassionate Response to a Medical Failure. In *Death, Dying and the Ending of Life, Volumes I and II*. Routledge, 2007.

Broome, John. "Is Rationality Normative?." *Disputatio* 2, no. 23 (2007): 1-18.

Broome, Matthew R. & Lisa Bortolotti. “Mental Illness As Mental: In Defense Of Psychological Realism.” *Humana Mente* 11 (2009): 25-44.

Camus, Albert. *The Myth of Sisyphus and Other Essay*s. New York: Alfred A. Knopf, 1955.

Card, Catherine. “How Facebook AI Helps Suicide Prevention.” Meta (blog), last modified 29 September 2021. https://about.fb.com/news/2018/09/inside-feed-suicide-prevention-and-ai/.

Celedonia, Karen L, Marcelo Corrales Compagnucci, Timo Minssen, and Michael Lowery Wilson. “Legal, Ethical, and Wider Implications of Suicide Risk Detection Systems in Social

Media Platforms." *Journal of Law and the Biosciences* 8, no. 1 (2021): 1-11. https://doi.org/10.1093/jlb/lsab021.

Centers for Disease Control and Prevention. "Promoting Individual, Family, and Community Connectedness to Prevent Suicidal Behavior." Atlanta, GA: Centers for Disease Control and Prevention (CDC), National Center for Inquiry Prevention and Control, 2008.

Chan, Kelvin. "TikTok Scrambles to Remove Suicide Video Clips, Ban Users." *AP News*, 29 April 2021. https://apnews.com/article/media-social-media-suicides-0cf6ea4e10ebe136beecbd079581337f.

Culver, C. M., and B. Gert. "Competence." In *The Philosophy of Psychiatry: A Companion*, edited by Jennifer Radden, 258-271. Oxford University Press, 2006.

David-Ferdon, Corinne, Alex E. Crosby, Eric D. Caine, Jarrod Hindman, Jerry Reed, and John Iskander. "CDC Grand Rounds: Preventing Suicide Through a Comprehensive Public Health Approach." *Morbidity and Mortality Weekly Report* 65, no. 34 (2016): 894–897.

D'Hotman, Daniel, and Erwin Loh. "AI Enabled Suicide Prediction Tools: A Qualitative Narrative Review." *BMJ Health & Care Informatics* 27, no. 3 (2020): 1-10. https://doi.org/10.1136/bmjhci-2020-100175.

D'Hotman, Daniel, Erwin Loh, and Julian Savulescu. "AI-Enabled Suicide Prediction Tools: Ethical Considerations for Medical Leaders." *BMJ Leader* 5, no. 2 (2021): 102–107. https://doi.org/10.1136/leader-2020-000275.

Dong, Min, Liang-Nan Zeng, Li Lu, Xiao-Hong Li, Gabor S. Ungvari, Chee H. Ng, Ines H. I. Chow, Ling Zhang, Yuan Zhou, and Yu-Tao Xiang. "Prevalence of Suicide Attempt in Individuals with Major Depressive Disorder: A Meta-analysis of Observational Surveys." *Psychological Medicine* 49, no. 10 (2019): 1691-1704. doi:10.1017/S0033291718002301.

Downie, Jocelyn. "Permitting Voluntary Euthanasia And Assisted Suicide: Law Reform Pathways For Common Law Jurisdictions." *QUT Law Review* 16, no. 1 (2016): 84-112.

Ebbott, Kristina. "A 'Good Death' Defined By Law: Comparing The Legality Of Aid-In-Dying Around The World." *William Mitchell Law Review* 37, no. 1 (2010): 172-205

Fischer, John M., and Mark Ravizza. *Responsibility And Control: A Theory Of Moral Responsibility*. New York: Cambridge UP, 1998.

Gomes de Andrade, Norberto Nuno, Dave Pawson, Dan Muriello, Lizzy Donahue, and Jennifer Guadagno. “Ethics and Artificial Intelligence: Suicide Prevention on Facebook.” *Philosophy & Technology* 31, no. 4 (2019): 669–684. https://doi.org/10.1007/s13347-018-0336-0.

Halsband, Aurélie, and Bert Heinrichs. 2022. “AI, Suicide Prevention and the Limits of Beneficence.” *Philosophy & Technology* 35, no. 4 (2022): 103. https://doi.org/10.1007/s13347-022-00599-z.

Horwitz, Allan V. “The DSM-5 and the Continuing Transformation of Normal Sadness Into Depressive Disorder.” *Emotion Review* 7, no. 3 (2015): 209-215.

Hoven, C.W., Mandell, D.J., & Bertolote, J.M. “Prevention Of Mental Ill-Health And Suicide: Public Health Perspectives.” *European Psychiatry* 25, no. 5 (2010): 252-256.

Kaste, Martin. “Facebook Increasingly Reliant on A.I. To Predict Suicide Risk.” *NPR*, 17 November 2018, sec. National. https://www.npr.org/2018/11/17/668408122/facebook-increasingly-reliant-on-a-i-to-predict-suicide-risk.

Kim, Kristen, and Jong-Ik Park. “Attitudes toward suicide among college students in South Korea and the United States.” *International Journal of Mental Health Systems* 8, no. 17 (2014): 1-5. doi: 10.1186/1752-4458-8-17

Langsam, Harold. "Rationality, Justification, and the Internalism/externalism Debate." *Erkenntnis* 68 (2008): 79-101.

“Instagram: New Tools to Ban Self-Harm and Suicide Posts.” *BBC News*, 11 November 2020. https://www.bbc.com/news/technology-54903428

James, Susan D. “Kipp Rusty Walker Kills Self on Oregon Stage.” *ABC News*, 18 April 2011. https://abcnews.go.com/Health/public-suicide-bend-oregon-horrifies-audience-now-seek/story?id=13403282

Kiesewetter, Benjamin. *The Normativity Of Rationality*. Oxford: Oxford University Press, 2017.

Lejeune, Alban, Aziliz Le Glaz, Pierre-Antoine Perron, Johan Sebti, Enrique Baca-Garcia, Michel Walter, Christophe Lemey, and Sofian Berrouiguet. “Artificial Intelligence and Suicide Prevention: A Systematic Review.” *European Psychiatry* 65, no. 1 (2022): E19. https://doi.org/10.1192/j.eurpsy.2022.8.

Lorenz, Taylor. “Internet ‘Algospeak’ Is Changing Our Language in Real Time, from ‘Nip Nops’ to ‘Le Dollar Bean.’” *Washington Post*, 11 April 2022. https://www.washingtonpost.com/technology/2022/04/08/algospeak-tiktok-le-dollar-bean/.

Luxton, David D., Jennifer D. June, and Jonathan M. Fairall. “Social Media And Suicide: A Public Health Perpsective.” *American Journal of Public Health* 102 (2012): S195-200.

Marchant, Amanda, Keith Hawton, Ann Stewart, Paul Montgomery, Vinod Singaravelu, Keith Lloyd, Nicola Purdy, Kate Daine, and Ann John. “A Systematic Review of the Relationship between Internet Use, Self-Harm and Suicidal Behaviour in Young People: The Good, the Bad and the Unknown.” *PloS One* 12, no. 8 (2017): e0181722. https://doi.org/10.1371/journal.pone.0181722.

Marks, Mason. “Artificial Intelligence Based Suicide Prediction.” *Yale Journal of Law and Technology* 21, no. 3 (2019). Available at SSRN: https://ssrn.com/abstract=3324874.

Public Health Agency of Canada. “Advance Contract Award Notice.” https://buyandsell.gc.ca/cds/public/2017/12/22/a6835bc7e03bc761b2d730ebf18fc33e/100019641 6_ai_acan_english.pdf

Meche, Monique. “Amplifying #SuicidePrevention Resources on Twitter.” Twitter (blog), 9 September 2020. https://blog.twitter.com/en_us/topics/company/2020/amplifying-suicideprevention-resources-on-twitter.

Memon, Aksha M., Shiva G. Sharma, Satyajit S. Mohite, and Shailesh Jain. “The Role of Online Social Networking on Deliberate Self-Harm and Suicidality in Adolescents: A Systematized Review of Literature.” *Indian Journal of Psychiatry* 60, no. 4 (2018): 384. https://doi.org/10.4103/psychiatry.IndianJPsychiatry_414_17.

Nelson, Lawrence and Erick Ramirez. “Can Suicide In The Elderly Be Rational?” In *Rational Suicide in the Elderly Clinical, Ethical, and Sociocultural Aspects,* edited by Robert E. McCue & Meera Balasubramaniam, 1-21. Springer, 2017.

Periyakoil, Vyjeyanthi S., Helena Kraemer, and Eric Neri. “Multi-Ethnic Attitudes Toward Physician-Assisted Death In California And Hawaii.” *Journal of Palliative Medicine* 19, no. 10 (2016): 1060-1065.

Pourmand, Ali, Jeffrey Roberson, Amy Caggiula, Natalia Monsalve, Murwarit Rahimi, and Vanessa Torres-Llenza. “Social Media and Suicide: A Review of Technology-Based

Epidemiology and Risk Assessment." *Telemedicine Journal and E-Health: The Official Journal of the American Telemedicine Association* 25, no. 10 (2019): 880–888. https://doi.org/10.1089/tmj.2018.0203.

Reuters Staff. 2019. "Facebook Bans Self-Harm Images in Fight Against Suicide." Reuters, 10 September 2019. https://www.reuters.com/article/us-facebook-content-idUSKCN1VV1XT

Rice, Simon, Jo Robinson, Sarah Bendall, Sarah Hetrick, Georgina Cox, Eleanor Bailey, John Gleeson, and Mario Alvarez-Jimenez. "Online and Social Media Suicide Prevention Interventions for Young People: A Focus on Implementation and Moderation." *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 25, no. 2 (2016): 80.

Sanburn, Josh. "A Brief History Of Self-Immolation." Time, 20 January 2011. https://content.time.com/time/world/article/0,8599,2043123,00.html

Selinger, Evan and Kyle Powys Whyte, Kyle Powys. "Competence And Trust In Choice Architecture." *Knowledge, Technology & Policy* 23 (2010): 461-482.

Snapchat. "Introducing Here For You." 9 August 2022. https://newsroom.snap.com/en-US/here-for-you.

Swanson Jeffrey W., Richard J. Bonnie, and Paul S. Appelbaum. "Getting Serious About Reducing Suicide: More 'How' And Less 'Why.'" *Journal of the American Medical Association* 314 (2015): 2229-2230. doi:10.1001/jama.2015.15566

Thaler, Richard H., and Cass R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. Penguin, 2009.

Wakefield, Jane. "Friend Challenges Facebook Over Ronnie McNutt Suicide Video." *British Broadcasting Service*, 20 September 2020. https://www.bbc.com/news/technology-54176205

Walsh, Colin G., Jessica D. Ribeiro, and Joseph C. Franklin. "Predicting Risk of Suicide Attempts Over Time Through Machine Learning." *Clinical Psychological Science* 5, no. 3 (2017): 457–469. https://doi.org/10.1177/2167702617691560.

Way, Jonathan. "The Normativity Of Rationality." *Philosophy Compass* 5, no. 12 (2010): 1057-1068.

Warnock, Caroline. “Ronnie McNutt: Facebook Live Stream Suicide Video Goes Viral On TikTok.” Heavy.Com, 7 September 2020. https://heavy.com/news/2020/09/ronnie-mcnutt-death-live-stream/

Weatherson, Brian. *Normative Externalism*. Oxford University Press, 2019.

World Health Organization. “*The Icd-11: International Statistical Classification of Diseases and Related Health Problems (ICD)*.” Geneva: World Health Organization, 2022.

World Health Organization. “*Preventing Suicide: A Global Imperative.”* Geneva, Switzerland. World Health Organization, 2016.

Young, Jerome. “Morals, Suicide, And Psychiatry: A View From Japan.” *Bioethics*, 16, no. 5 (2002): 412-424.