The ethical role of computational linguistics in digital psychological formulation and suicide prevention

Martin P. Orr
Auckland University of Technology
`martinorr521@gmail.com`

Kirsten van Kessel
Auckland University of Technology
`kirsten.vankessel@aut.ac.nz`

David Parry
Murdoch University
`david.parry@murdoch.edu.au`

Abstract

Formulation is central to clinical practice. Formulation has a factor weighing, pattern recognition and explanatory hypothesis modelling focus. Formulation attempts to make sense of why a person presents in a certain state at a certain time and context, and how that state may be best managed to enhance mental health, safety and optimal change. Inherent to the clinical need for formulation is an appreciation of the complexities, uncertainty and limits of applying theoretical concepts and symptom, diagnostic and risk categories to human experience; or attaching meaning or weight to any particular factor in an individual's history or mental state without considering the broader biopsychosocial and cultural context. With specific reference to suicide prevention, this paper considers the need and potential for the computational linguistics community to be both cognisant of and ethically contribute to the clinical formulation process.

1 Introduction

Modelling is central to mental healthcare. Deficits in modelling, or failure to understand and manage those deficits, can lead to deficits in care.

Risk prediction, the diagnostic process, and key phenomena identification and monitoring such as mood symptoms are valid targets for the application of computational linguistics to suicide prevention. However, from a clinical perspective each of these targets and the research and categorical conceptual modelling that underlie them has major limitations, complexity, and contention (Chakraborty, 2020; Franklin et al., 2017; Fried, 2015; Large, 2018; Turner et al., 2021; Waszczuk et al., 2017).

Many aspects of mental health clinical practice are based on limited theoretical models, limited data and limited resources and involve varying presentations, preferences and levels of understanding, strengths, insight, and engagement. Formulation is the key clinical process for attempting to integrate these multiple interacting limited models and factors, to create an overall working model on which to base future action and interventions (de Beer, 2017; Carey and Pilgrim, 2010; Challoner and Papayianni, 2018).

Clinical formulation has a pattern recognition, factor weighing and explanatory hypothesis modelling focus. Formulation attempts to make sense of why a person presents in a certain state at a certain time and context and how given the known vulnerabilities, strengths, preferences and available resources that state may be best changed in a safe and effective way (Critchfield et al., 2022; Fernando et al., 2012; Johnstone and Dallos, 2013; Mace and Binyon, 2005; Manjunatha, 2019).

In keeping with the evolution and variation of mental health practice, formulation has historically taken varying forms and had varying drivers, and had questions raised about its validity and utility. However formulation retains a central role in care delivery, is considered as requiring the highest level of clinical expertise and is a key component of examination for specialist qualification (de Beer, 2017; Challoner and Papayianni, 2018; Sullivan et al., 2020).

Inherent to the clinical need for formulation is an appreciation of the complexities, uncertainty and limits of applying categories and theoretical concepts to human experience or attempting to attach meaning or weight to any particular factor in an individual's history or mental state without considering the broader context.

This paper considers the opportunities and challenges for computational linguistics in emulating and augmenting the clinical formulation process and contributing to broader related digital mental health developments. Highlighted is the need to appreciate the ethical and clinical safety risks, particularly if developments in the computational linguistics field are misperceived or exaggerated in terms of their certainty and capacity for suicide

prediction and reduction.

The paper discusses the clinical assessment and planning process and the phenomenological psychopathology analysis, nosological diagnostic classification, individual psychodynamics and risk prediction complexities that drive the need for formulation. The concepts of mood, affect and emotion are discussed to illustrate some of the issues around the standardised interpretation of human experience and classification into diagnoses. The ethics and difficulties of attempting to predict or modify the risk of low base rate complex emergent events such as suicide is highlighted (Woodford et al., 2019; World Health Organization, 2014). A structure for formulation is provided to highlight the key components and where computational linguistics may be of assistance.

The central arguments will be that the data gathering, pattern recognition, factor weighing, and modelling of clinical formulation are areas in which computational linguistics could and should assist. Pattern recognition and modelling around words and language in context is central to mental health clinical practice and computational linguistics. Mental health and computational linguistics specialists can synergically use language as a method to gain insight and formulate a model of another's consciousness, intent and experience. This can contribute to risk, diagnostic and psychodynamic formulation. However, appreciation of the limitations of modelling and prediction particularly in application to suicide prevention will remain central. The Artificial Intelligence (AI) ethical principles of autonomy, justice, beneficence non-maleficence and explicability will remain a challenge and a duty for the CLPsych community (Floridi and Cowls, 2019). Appreciating the rationale for the utilisation of formulation in clinical practice and seeking to place the ethos and process of formulation at the heart of computational linguistics practice to enhance explicability will assist in addressing that duty. Machine learning and computational linguistics may play a role in more accurately identifying the contextual and contingent factors and the level of certainty or uncertainty inherent in the formulation modelling and explanatory hypothesis.

The primary purpose of this work is to provoke thought and facilitate further conceptual and operational ethical co-design of digital formulation. The aim is to help build a shared understanding

of the rationale, structure and process of clinical formulation and call upon the CLPsych community to consider what contributions they could make to digitally enable and improve it particularly within a suicide prevention context. It is recognized there is a concept-reality gap between what clinicians might ideally desire and what the computational linguistic field is currently able to offer (Orr and Sankaran, 2007). However, the CLPsych community could play an important role in clarifying and developing the conceptual vision for digital formulation, and the required technological and methodological steps to get there.

The paper is intentionally largely technology and data source agnostic and focused on the clinical need and related medicolegal and ethical principles. The aim is to stimulate rather than limit thought or argue for a particular technological or methodological direction. The paper will touch on the initial steps to cross the concept-reality gap the authors are taking. This includes a focus on ethics, digital transformation, sleep and suicide, social media data and integrated thematic analysis and topic modelling.

2 The role and place of formulation in clinical practice

This next section aims to briefly set out some key concepts on which to build a shared understanding of the need for and place of formulation in clinical practice particularly in suicide prevention. These concepts are complex and contentious with differing definitions and scopes and varying degrees of clinical understanding and application in practice. Highlighted are the roles and limitations of language, phenomenology, nosology and risk prediction.

3 Language as a window into mental and brain state

There is limited understanding of the nature of consciousness or the mind and how this relates to brain function (Frith, 2021; Graziano, 2021). However, there is a general understanding that integrated biological, psychological and sociological factors impact on brain function and impact on the integrated experience and expression of thoughts, emotion, and behaviour. Machine learning affords the capacity to dynamically identify and analyse multiple signals indicative of an individual's mental state and intent. These signals may be neurophys-

iological, behavioural and of increasing interest to suicide prevention natural language, including that occurring in social media (Resnik et al., 2020; Chancellor and De Choudhury, 2020; Coppersmith et al., 2018; Fonseka et al., 2019).

To gain a greater timely understanding of the lived experience and meaning of suicidal thoughts and behaviour we need a greater appreciation of the dynamic cognitions and emotion and contexts that colour an individual's thoughts and drive them to action (Harris and Barraclough, 1997; Liu et al., 2020; Marsh, 2018). Social media data may provide an additional window and insights into this experience and an opportunity to intervene in a timely way.

Clinically language is a key tool for assessing and communicating thoughts, emotion and behaviour. Language is central to the assessment of mental state and from this potential brain state. Language assists in making hypotheses about electrochemical and cognitive processes in a section or circuit of the brain at a particular point in time that hence drive physical, biological, psychopharmacological and psychosocial interventions.

Language is a significant window into human experience but may not always provide an accurately drawn picture of reality. The image may be skewed and distorted by faulty mental models, cognitive biases, and misinterpretations by both the experiencer and the observer. Computational linguistics as the study of language using computational methods and theoretical models, similarly to clinical practice, has an inherent interest in ensuring any model deficits or conflicts are understood, minimised and managed.

4 Phenomenological psychopathology and nosology

Phenomenological psychopathological analysis is the process that underlies the clinical perception and interpretation of the experience and behaviour of others (Aftab and Ryznar, 2021; Chakraborty, 2020; Nelson et al., 2021).

Nosology is the classification of medical diseases. Nosological modelling can occur at three levels: aetiological (disease cause is known) pathogenetic (disease process is known) and symptom (only reported or interpreted experience is known). Mental disorder diagnoses are typically at the symptom modelling syndrome level (Kendler, 2009; Aftab and Ryznar, 2021).

Human experience and behaviour are characterised by a dimensional nature and multifactorial temporal contextual determinants. Complexity, and ambiguity is inherent. There is only limited knowledge of the causes and mechanisms by which mental disorders and perceived aberrant experience and behaviour arise. Accordingly, there are only theoretical models of varying fidelity and evidence base and agreement around the nature and classification of mental disorder, and how experience and behaviour should be interpreted and determined to be pathological. Similarly, the selection and mechanism of action of interventions, their benefits and harms, and predictions and determinants of prognosis all require the interpretation and weighing of various population research models as to what may be best and available for a specific patient in a specific mental state, in a specific time and context.

Risk categories, and diagnostic categories based on the identification and interpretation of phenomena and syndromes have significant reliability, validity and intervention, prognostic and safety limitations (Michelini et al., 2021; Nelson et al., 2021).

Although the terms affect, emotion and mood are often used interchangeably, they have a broad historical range of interrelated but separate specific meanings, definitions and perceived implications arising from variations (Berrios, 1985).

An emotion can be understood as the subjective personal experience and interpretation of a feeling state. Affect refers to an assessor's interpretation of the emotional experience of another, and typically includes not just reference to the type, but also the range and stability and appropriateness of expressed emotion within a specific context.

Emotions may be of short duration and fluctuate and represent the subjective interpretation of chemically induced physiological experience. The interpretation of this physiological emotional experience may be influenced by the longer standing and more prominent mood state, which may have a complex biopsychosocial basis.

A report of a mood symptom such as depression may have significant differing impact, relevance and meaning depending on the pattern intensity, duration, associations and context of occurrence. It may be a sign of a brief adjustment to a stressor, an indication of emotional dysregulation in someone with a personality disorder, form part of various levels and presentations of a major depressive disorder, be associated with medical and neurological

disorders from dementia to Parkinson's disease, be associated with or secondary to drug use prescribed and illicit, form part of a broader bipolar disorder, or be an early presentation or association with schizophrenia. Weight may be given to one diagnosis over another if there is a clear family or personal history or pattern of a particular disorder and other known risk, symptom, sign and contextual factors are present.

Diagnoses can be of use in care planning, funding, research and making predictions about the future. However, they have significant limitations, not least if it is forgotten they are syndromal level models, that tell little of the personal story and context of the individual. The symptom and sign and temporal components of the diagnostic model may be subject to deficits or non-standardisation in interpretation and report. Race, culture, gender, age, language, education, intellectual and sensory impairment and economic status and societal marginalisation may all have an impact on the expression and interpretation of experience and behaviour. These factors may contribute to significant inter-rater variability as to what diagnosis or diagnoses are ascribed to an individual. Individuals that receive a specific mental disorder diagnosis, may have significant variation in terms of what criteria they meet, their individual experience, and underlying causal and mechanism of development factors. For example, in the DSM classification system 227 combinations of criteria can lead to a diagnosis of major depression, including 64 combinations which don't require a report of depressed mood. Some combinations may be more common and more meaningful from a clinical priority and potential to intervene perspective and computational linguistics could assist with identifying these (Zimmerman et al., 2015).

5   Suicide risk prevention

There are major challenges, limitations and clinical and ethical risks in trying to predict complex multi-factorial emergent low base rate events such as suicide that have a high magnitude of adverse consequence if that prediction is wrong (Pridmore, 2015; Nock et al., 2019; Large et al., 2017). The majority of those classified as being at high risk of suicide, do not commit suicide, and the majority of suicides will emerge from those classified as low risk, or who have not been assessed for suicide risk, have not expressed suicidal ideation or whom

are not engaged in services (McHugh et al., 2019; Kessler et al., 2020; Durie, 2017; Large, 2018). The expression and actioning of suicidal intent can vary in intensity and fluctuate rapidly and be influenced by ambivalence, mood and emotional state, change in perceived circumstances, level of trust, wish to protect others, shame, denial, rationalisation, coping patterns, cognitive impairment, gain, and impulsivity (Yaseen et al., 2019; Galynker et al., 2017; Deisenhammer et al., 2009; Freedenthal, 2007).

There is a need to appreciate that even if increase the specificity and sensitivity of a technology capable of screening for a particular disorder, behaviour or risk, the positive predictive and negative predictive value will vary as a factor of prevalence in the targeted community. Suicide is a low base rate event making prediction complex and making the capacity for undue harm and intrusive unnecessary interventions higher. Even if the sensitivity and specificity of a test for suicide is significantly improved the positive predictive value may still be relatively low. This is not an argument to stop researching computational linguistics' capacity to improve suicide prediction but is a call to be cognisant of the limitations and to take a broader view on how machine learning and computational linguistics may contribute to suicide risk management.

Different people will have different pathways, processes, contexts, and timelines that take them to suicide. Some may have a more linear escalating suicide risk chain they follow; others will display a more complex emergence pattern where multiple factors came together at a certain point in time and a chain is only apparent with retrospective coherence (Kurtz and Snowden, 2003).

While making suicide predictions has inherent complexity and limitation, enhancing the capacity for machine learning to detect risk signals and offer support to reach out and seek help, would provide a chance to positively change that pathway and context (Tielman et al., 2019; Ryan, 2015). Machine learning may be able to assist in identifying what key potential contextual risk factors are for an individual or community; assist in the triage and prioritisation of attention for that individual or community; and do this at a speed and scale over multiple sources that exceeds human capabilities (Resnik et al., 2020; Shing et al., 2020).

Clinical risk including suicide risk needs to be considered in relation to a specific population and in relation to the individual's own baseline or typ-
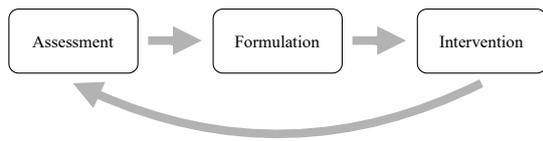
Figure 1: Central Role of Formulation

ical risk level or pattern. Timing and context is important including the presence or absence of key precipitants or protective factors. Ascertaining what factors are able to be managed, minimised or developed, and which resources are ideally and in reality, available are also important concerns. Rather than just provide a simple risk category, formulation aims to provide an integrated weighted contextualised view of all factors that may have a role in intervention planning and risk management (Pisani et al., 2016; Wyder et al., 2021; Fitzpatrick, 2018; Kessler et al., 2020).

6 Formulation at the centre of clinical practice

The clinical assessment and intervention process typically involves the key dynamic, integrated, iterative stages of history taking, mental state examination, formulation, diagnosis and care planning. Formulation is at the centre, prioritising and integrating key aspects of the assessment as a foundation for the personalised intervention planning.

Formulation can be perceived as a form of clinical storytelling. Clinical formulation includes the recurrent patterns, key themes, plot points and relationships, and cultural and contextual factors that characterise and help draw a mental model of an individual and their world.

Many individuals even if they have never had a formal mental health diagnosis before, may be found on assessment to have had recurrent patterns suggestive of previous episodes or prodrome or vulnerabilities. Those that have an established recurrent relapsing disorder, may have patterns of risk behaviours and contexts and early warning signs, that the client has varying and fluctuating levels of insight into, but that may be well recognised by families and supports.

Storytelling in written and spoken language has traditionally been a way to transmit knowledge and understanding to a group and through generations. The narrative structure of storytelling may assist in human recall and motivational understanding. Clinical formulation is a structured way to make

sense and meaning of another's consciousness and experience and convey the story of their life, with a view to positively influencing the next stage of that story. Multiple factors and models are weighed and weaved using a structured process of analysis and reporting. Computational linguistics with its strengths in factor identification, weighing and integrated pattern recognition across multiple sources and contexts could assist this process.

Unlike describing the precise formulation or composition of a chemical compound or drug, the specific elements or processes by which a human experience arises is unknown. There is limited knowledge of the aetiology or pathogenesis of mental disorders, nor the nature of consciousness or emotions or experience. However, there are a range of biological, psychological and sociological research-based models that may guide the formulation process. The quality of the formulation is dependent on the knowledge and skill of the clinician in history taking and mental state examination and being able to identify key patterns, vulnerabilities, strengths, relationships and structures and integrate the findings with appropriate theoretical models. The quality of the formulation can be iteratively improved by the availability of additional data sources and the input of the multiple stakeholders, family, supports and caregivers that may play a role in an individual's life (Ford et al., 2019; Geach et al., 2018; Johnstone, 2018).

7 Clinical formulation structure

The clinical formulation may be approached in a structured manner, with data and key findings captured under a series of interrelated headings, each capturing a different but interrelated descriptive, theoretical or explanatory perspective (Chang and Lundahl, 2019; Weerasekera, 1993). Machine learning and computational linguistics operating under various levels of autonomy, could assist in augmenting this clinical pattern recognition and modelling process.

Some variant of a series of "P" headings such as problem, predisposing, precipitating, perpetuating, protective factors and prognosis headings is common clinical practice. Patterns, preferences, and priorities have been added for this work to emphasise these key attributes of the formulation process and where machine learning and computational linguistics could play a key role in capturing and weighing and providing decision support.

Problem	What are the key findings from the presenting complaint and history of presenting complaint, and the mental state examination, that characterize the problem or disorder?
Predisposing	What biopsychosocial and cultural contextual factors may have predisposed the individual to the disorder?
Precipitating	What biopsychosocial and cultural contextual factors may have precipitated the problem or exacerbation of the disorder?
Perpetuating	What biopsychosocial and cultural contextual factors may perpetuate or exacerbate the problem or provide a barrier to recovery?
Protective	What are the biopsychosocial and cultural contextual factors that may offer protection, assist in recovery or prevent further harm or adverse outcomes?
Prognosis	What is the expected response and outcome for this individual given what is known from population research and their specific history and level of insight, impairment, vulnerabilities, strengths and resources? How might interventions work or not work or cause harm and in what context and time?
Patterns	What patterns may be evident in the history and how may these relate to known psychological models?
Preferences	How does the individual prefer to understand or model their problem(s) for themselves and what resources and interventions do they prefer to utilize and how may these preferences be impacted on by insight and judgement?
Priorities	What are the priorities for the intervention plan given the knowledge about the individual, their past response and preferences, available resources and logistics, local and professional best practice guidelines, and relevant science?

Table 1: Clinical Formulation Structure

By considering biological, psychological and social facets and individual and systemic contextual components of each heading, the key formulation issues are often captured in a biopsychosocial grid structure, before being converted into an integrated coherent written form (Weerasekera, 1993). Some structures consider culture as inherent to the biopsychosocial analysis; others draw it out as a separate heading or separate cultural formulation process to ensure this important factor is focused on and not neglected.

Conceptual models created from international data may have transcultural limitations. This may be a significant issue when conceptual models are being utilised in formulation and particularly when involving the interpretation of experience and behaviour. DSM5 diagnoses are essentially conceptual models built on international data that contribute to care by providing a framework to describe an individual's perceived experience. However, the framework may negatively impact on care if the transcultural limitations are not adequately addressed or understood (Bredström, 2019; Rangihuna et al., 2018; La Roche et al., 2015).

Similarly, computational linguistics research and application in the mental health and suicide prevention domains needs to be designed and interpreted with a sociocultural contextual awareness as is emphasised by the formulation ethos (Durie, 2017; Hatcher et al., 2017; Lawson-Te Aho Dr, 2017; McClintock and McClintock, 2017).

8 Ethical and regulatory issues

The following section outlines a range of ethical and regulatory issues that are important considerations when developing and deploying digital mental health interventions particularly in the area of suicide prevention. Health interventions should be evidence based, and subject to academic, clinical governance, regulatory and ethical review. There is a need to be ethically cognisant of the risk-benefit profile, the relative utility and costs and the numbers of people that may additionally benefit or be harmed by an intervention within a specific context. The relevance, meaning, sensitivity and specificity of screening and diagnostic tests must be described with reference to a stipulated time period, prevalence and clinical context (Andrade, 2015). Digital mental health interventions including computational linguistic based suicide prevention interventions need to be subject to similar standards.

Primum non nocere or "first, do no harm" is a fundamental principle of bioethics. Failure to understand the complexity and limitations of suicide risk prediction has significant capacity to cause harm. Simplistic, generalised or static risk categorisation can lead to unintended harm and there is a need for dynamic formulation based assessment that recognises the importance of context for an individual's strengths and vulnerabilities.

The analytic power, reach, personalisation, timeliness and vigilance of AI based digital care affords major potential benefit. However, AI can be intrusive, discriminative, unwanted, and wrong. In suicide prevention resources could be allocated to the wrong groups, to the wrong individual, or be of the wrong type or quality and quantity. Some individuals may have unnecessary protections or intrusions placed on their lives which are damaging or disabling (McKernan et al., 2018). AI algorithms are subject to bias, misuse, undue trust, and unintended consequences and require continuous ethically based and sociocultural aware research, co-design, and governance (Yu, 2020; Floridi et al., 2020; Stein and Reed, 2019; Challen et al., 2019).

Continually striving to improve AI based risk prediction and management at an individual to societal level and getting to zero people dying by suicide is a morally worthy goal. However, there is a need to consider how the nature and current status of attaining that goal may be societally interpreted or misinterpreted. Stigma can have an impact on suicide bereavement and is an important consideration for suicide postvention. Bereaved family and caregivers can experience significant stigma, shame and blame and societal judgement based on a belief that they should have seen the signals of pending suicide and predicted and prevented the death (Evans and Abrahamson, 2020). In the reporting of improvements in suicide risk prediction, it is important that the CLPsych community highlight the ongoing complexities and limitations in identifying, seeing, analysing and acting on the signals and do not unintentionally contribute to exacerbating suicide bereavement and stigma.

If a digital system claims a clinical or therapeutic intervention function, then the system can be expected to be held to a high ethical and regulatory standard. This includes requiring a high level of mandated understanding of how the system integrates into broader clinical care processes, medicolegal responsibility and governance frameworks and whether it potentially requires software as a medical device type certification. There are varying developing regulatory standards and definitions for medical device type software. An AI based system providing triage and treatment advice where an individual may be at risk of suicide would likely present some of the highest ethical and clinical risks for development and deployment in a healthcare context and attract the highest regulatory categorisation and governance requirements (NEAC, 2019; Fernandes and Chaltikyan, 2020; Keutzer and Simonsson, 2020).

There is increasing interest in the use of social media and AI in suicide prevention. The international literature on social media research highlights various contentions including defining public vs private data, consent and anonymity and minimising bias and algorithmic harm (Townsend and Wallace, 2016; Chiauzzi and Wicks, 2019). There is increasing recognition of a need for social media-based research to have ethical overview to ensure that quality research is being proposed that understands the limitations and context of the data analysis and is protective and respectful of potentially vulnerable communities (British Psychological Society, 2017; Townsend and Wallace, 2016; Pagoto and Nebeker, 2019; Chiauzzi and Wicks, 2019; Benton et al., 2017).

Tutelary law and ethics, relates to those aspects of the legal and ethical system that have a focus on guardianship and protection (Unsworth, 1991). Mental healthcare services have had a long and difficult history with care and protection and guardian roles. The legal system is aware that good protective intents do not always result in good or optimal outcomes and there is always a need to consider who will guard the guardians. Clinical decisions and opinions about risk that impact on an individual's civil liberties, are often subject to review by tutelary mental health courts and tribunals; decisions influenced by AI based categorisation or predictions should similarly be expected to be reviewed by the tutelary system (Szmukler, 2014).

Floridi and Cowls (2019) have argued AI ethics can be reduced to five core principles. Four of these are the traditional bioethical principles of autonomy, justice, beneficence and non-maleficence to which they have added explicability. Explicability aims to capture the concepts of intelligibility and accountability. Building suicide prevention interventions and research on faulty, limited or poorly

understood or described models affords significant ethical and clinical risk. Clinicians and researchers need to take a lead in ethically shaping and governing the emergent capacity for greater levels of social media and AI based suicide prevention research and development(Hom et al., 2017; Hunter et al., 2018; Pagoto and Nebeker, 2019). Before the deployment of AI in a mental health setting, stakeholders should have an adequate understanding of how it was co-designed and works and who is accountable and liable for how it works (Floridi and Cowls, 2019; Price et al., 2019). The formulation process could improve explicability in that there should be a clearer, intelligible and accountable process as to why intervention decisions were made. This should include having an understanding of the mental health theoretical models on which or for what, the machine learning algorithms were built (de Andrade et al., 2018).

9 Discussion and conclusion

In clinical practice there are significant standardisation, ethical, safety and effectiveness issues when classifying an individual as in or out of some binary diagnostic or risk category. This is particularly so when the constructs or models that underlie each criterion are limited in their scientific basis and are not operationally defined and there is significant variation in training, interpretation and application and perceived clinical utility.

Similarly, when computational linguistics developments aim to assist in symptom, diagnostic and suicide risk prediction categorisation there may be significant theoretical, ethical, utility and clinical safety concerns and limitations. There is a need to move beyond risk categorisation to risk formulation as part of the broader clinical formulation and intervention context. Any risk prediction categorisation produced needs to be treated like the output from any screening or diagnostic test; that is as another datapoint for the formulation, that is to be iteratively weighed, integrated and interpreted within the broader dynamic clinical context and not considered definitive or static.

Human experience is often time and context dependent, dynamic and multidimensional and occurs along a spectrum rather than within discrete categories. Formulation is the key focus of natural clinical intelligence and ought to be a key focus for artificial intelligence.

Computational linguistics could help in the development and assessment of a broader contextual understanding of an individual's history and mental state. A diagnostic and risk formulation process affords the opportunity to present a richer personalised explanatory model that links all the factors and highlights the complexity, uncertainty and importance of context and dynamic change.

The clinical formulation process of iterative factor identification and weighing, pattern recognition and modelling, is in keeping with the strengths of machine learning and the computational linguistics process. There are opportunities for significant synergy. Computational linguistics can operate at a speed and scale of factor identification and analysis across multiple sources beyond human capability. The machine learning process may be refined on previous clinical assessments, with emphasis given to mental state examinations and formulations. Clinician in the loop training and curation processes may assist with explicable and reflexive algorithmic improvement and production of meaningful safe ethical outputs. Though such formal clinical data may be difficult to access for current researchers, this can be expected to improve as machine learning is integrated and normalised as part of care delivery.

Machine learning and computational linguistics could improve the explicable quality of the acquisition, analysis and description of formulation data. Machine learning could also improve the quality of the theoretical models applied by improving the quality of research that underlies those models. There may be different and changing reasons and typologies for suicide and AI enabled research may be able to better timely categorise, trend and define these at an individual and community level (Clapperton et al., 2020; Martin et al., 2020).

In formulation every current and emergent finding needs to be iteratively analysed in context, and with knowledge of the strengths and limitations of the related clinical theoretical models. Particularly in suicide risk management there is a need to be highly cognisant of the difficulty and contention of predicting complex low base rate events and the harm that can result from both false negatives and false positives.

Digital transformation and co-design in AI empowered suicide prevention requires the working together of clinicians, communities, consumers, and digital media companies. The leverage, reach and analysis of AI empowered digital media make

taking a co-design and societal perspective more meaningful and achievable and anything more limited, less ethically justifiable.

Looking to the future computational linguistics could assist in the creation of a self-constructing and updating digital formulation drawing on multiple sources from social media to email to clinical notes and assessments to conducting autonomous interviews in oral and written format. These services could be delivered in the form of customisable digital guardians, coaches or clinicians that address, with varying levels of expertise, medicolegal responsibility and autonomy, the assessment, formulation and intervention process.

In terms of an example of potential next research steps the authors are currently integrating qualitative thematic analysis with machine learning based topic modelling to study sleep related concerns in a large social media based suicidality dataset. Sleep disturbances from insomnia to nightmares to sleep disordered breathing are associated with an increased risk of suicidal behaviour and night-time is a high-risk period for suicide (Braun and Clarke, 2006, 2019; Blei et al., 2003; Blei, 2012; Fast et al., 2016; Shing et al., 2018, 2020; Zirikly et al., 2019; Porras-Segovia et al., 2019; Tubbs et al., 2019). The research is exploring whether this integrated thematic analysis and topic modelling approach can contribute to the development of an explicable conceptual linguistic sleep signal model for AI empowered clinical formulation, prioritisation, treatment category recommendation, and psychoeducation in the area of suicide prevention. Identifying key topics and themes and a related lexicon are central to these clinical processes. Suicide is complex and multifactorial. By focusing on one potential signal (sleep), one machine learning technique (topic modelling) and one dataset the aim is a greater conceptual understanding of the opportunities and challenges that could be presented by an multi-signal, multi-source, explicable AI and formulation based suicide prevention system. The current focus is on social media data, but a range of biopsychosocial data sources might be integrated into a future system. Formulation and broader data contained in clinical assessments and discharge summaries could be a future key target for both analysis and enrichment (Adnan et al., 2013).

Developments in digital formulation from a clinician augmentation or decision support role to a more autonomous social media focused digital guardian, coach or clinician role will have major clinical and societal impact. Suicide is a complex time and context dependent phenomenon. There is increasing recognition of the need to broaden the clinical service focus of suicide prevention to a more societal level focus that has more timely vigilance and leverages a greater range of resources.

Clinic-based services accessing social media data, and social media based services accessing formal clinical data, and the integration of such services raises significant medicolegal, security and ethical issues (Williams et al., 2017; Price et al., 2019; Bhatia-Lin et al., 2019). However, machine learning assisting in the expansion from a clinical service to societal focus allows for more protective layers and opportunities for integrated formulation and intervention at more time points. Social media machine learning based interventions have the advantage that even if they have only a relatively small effect size on reducing suicidal behaviour they can be deployed at such scale and minimal marginal cost that they may have a significant impact at a population societal level (Torok et al., 2020).

Socialising emergent concepts, among research and practice leaders, is an important stage in the innovation diffusion and health practice change process (Beausoleil, 2018; Taherdoost, 2018; Rahimi et al., 2018). This can lead to critical analysis of relative advantage, adoption challenges and health impact , and feed through to strategic research, implementation and governance plans(Renken and Heeks, 2019). This paper has aimed to socialise the concept of digital psychological formulation with the goal of making a positive health impact on suicide prevention by promoting adoption and development of the concept by the CLPsych community.

There is a significant concept reality gap between current developments and getting to a stage of digital formulation being utilised by human and digital clinicians as part of standard mental health and suicide prevention practice (Heeks, 2006; Orr and Sankaran, 2007). However, it is a concept reality gap that is potentially fast narrowing and that the CLPsych community has both the developing expertise and ethical duty to take a leadership role in crossing, in a safe and clinically effective manner.

References

Mehnaz Adnan, Jim Warren, and Martin Orr. 2013. Semlink—dynamic generation of hyperlinks to enhance patient readability of discharge summaries. In *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*, pages 35–40. IEEE.

Awais Aftab and Elizabeth Ryznar. 2021. Conceptual and historical evolution of psychiatric nosology. *International Review of Psychiatry*, 33(5):486–499.

Chittaranjan Andrade. 2015. The numbers needed to treat and harm (nnt, nnh) statistics: what they tell us and what they do not. *The Journal of clinical psychiatry*, 76(3):12971.

Angele Marie Beausoleil. 2018. Revisiting rogers: the diffusion of his innovation development process as a normative framework for innovation managers, students and scholars. *Journal of Innovation Management*, 6(4):73–97.

Adrian Benton, Glen Coppersmith, and Mark Dredze. 2017. Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing*, pages 94–102.

German E Berrios. 1985. The psychopathology of affectivity: conceptual and historical aspects. *Psychological medicine*, 15(4):745–758.

Ananya Bhatia-Lin, Alexandra Boon-Dooley, Michelle K Roberts, Caroline Pronai, Dylan Fisher, Lea Parker, Allison Engstrom, Leah Ingraham, and Doyanne Darnell. 2019. Ethical and regulatory considerations for using social media platforms to locate and track research participants. *The American Journal of Bioethics*, 19(6):47–61.

David M Blei. 2012. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101.

Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4):589–597.

Anna Bredström. 2019. Culture and context in mental health diagnosing: Scrutinizing the dsm-5 revision. *Journal of Medical Humanities*, 40(3):347–363.

British Psychological Society. 2017. Ethics guidelines for internet-mediated research. *Leicester, UK: British Psychological Society*.

Timothy A Carey and David Pilgrim. 2010. Diagnosis and formulation: What should we tell the students? *Clinical Psychology & Psychotherapy*, 17(6):447–454.

Nandini Chakraborty. 2020. The importance of embedding psychopathology and phenomenology in clinical practice and training in psychiatry. *BJPsych Advances*, 26(5):287–295.

Robert Challen, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards, and Krasimira Tsaneva-Atanasova. 2019. Artificial intelligence, bias and clinical safety. *BMJ Quality & Safety*, 28(3):231–237.

Harriet Challoner and Fani Papayianni. 2018. Evaluating the role of formulation in counselling psychology: A systematic literature review. *The European Journal of Counselling Psychology*, 7(1).

S. Chancellor and M. De Choudhury. 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ Digit Med*, 3(1):43.

Anita Kumar Chang and Leslie H Lundahl. 2019. Which problem are we addressing today? the utility of a multifaceted formulation approach to a complex case. *American Journal of Psychotherapy*, 72(1):29–33.

E. Chiauzzi and P. Wicks. 2019. Digital trespass: Ethical and terms-of-use violations by researchers accessing data from an online patient community. *J Med Internet Res*, 21(2):e11985.

A. Clapperton, L. Bugeja, S. Newstead, and J. Pirkis. 2020. Identifying typologies of persons who died by suicide: Characterizing suicide in victoria, australia. *Arch Suicide Res*, 24(1):18–33.

Glen Coppersmith, Ryan Leary, Patrick Crutchley, and Alex Fine. 2018. Natural language processing of social media as screening for suicide risk. *Biomedical informatics insights*, 10:1178222618792860.

Kenneth L Critchfield, Francesco Gazzillo, and Ueli Kramer. 2022. Case formulation of interpersonal patterns and its impact on the therapeutic process: Introduction to the issue. *Journal of Clinical Psychology*.

Norberto Nuno Gomes de Andrade, Dave Pawson, Dan Muriello, Lizzy Donahue, and Jennifer Guadagno. 2018. Ethics and artificial intelligence: suicide prevention on facebook. *Philosophy & Technology*, 31(4):669–684.

Wayne A de Beer. 2017. Original opinion: the use of bloom's taxonomy to teach and assess the skill of the psychiatric formulation during vocational training. *Australasian Psychiatry*, 25(5):514–519.

Eberhard A Deisenhammer, Chy-Meng Ing, Robert Strauss, Georg Kemmler, Hartmann Hinterhuber, and

Elisabeth M Weiss. 2009. The duration of the suicidal process: how much time is left for intervention between consideration and accomplishment of a suicide attempt? *Journal of Clinical Psychiatry*, 70(1):19.

Mason Durie. 2017. Indigenous suicide: the turamarama declaration. *Journal of Indigenous Wellbeing*, 2(2):59–67.

Amy Evans and Kathleen Abrahamson. 2020. The influence of stigma on suicide bereavement: A systematic review. *Journal of Psychosocial Nursing and Mental Health Services*, 58(4):21–27.

Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 4647–4657.

Fara Aninha Fernandes and Georgi V Chaltikyan. 2020. Analysis of legal and regulatory frameworks in digital health: A comparison of guidelines and approaches in the european union and united states. *Journal of the International Society for Telemedicine and eHealth*, 8:e11 (1–13).

Irosh Fernando, Martin Cohen, and Frans Henskens. 2012. Pattern-based formulation: a methodology for psychiatric case formulation. *Australasian Psychiatry*, 20(2):121–126.

Scott J Fitzpatrick. 2018. Reshaping the ethics of suicide prevention: responsibility, inequality and action on the social determinants of suicide. *Public Health Ethics*, 11(2):179–190.

Luciano Floridi and Josh Cowls. 2019. A unified framework of five principles for ai in society. *Harvard Data Science Review*.

Luciano Floridi, Josh Cowls, Thomas C King, and Mariarosaria Taddeo. 2020. How to design ai for social good: Seven essential factors. *Science and Engineering Ethics*, 26(3):1771–1796.

Trehani M Fonseka, Venkat Bhat, and Sidney H Kennedy. 2019. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Australian & New Zealand Journal of Psychiatry*, 53(10):954–964.

E. Ford, K. Curlewis, A. Wongkoblap, and V. Curcin. 2019. Public opinions on using social media content to identify users with depression and target mental health care advertising: Mixed methods survey. *JMIR Ment Health*, 6(11):e12942.

Joseph C Franklin, Jessica D Ribeiro, Kathryn R Fox, Kate H Bentley, Evan M Kleiman, Xieyining Huang, Katherine M Musacchio, Adam C Jaroszewski, Bernard P Chang, and Matthew K Nock. 2017. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. *Psychological bulletin*, 143(2):187.

Stacey Freedenthal. 2007. Challenges in assessing intent to die: can suicide attempters be trusted? *OMEGA-Journal of death and dying*, 55(1):57–70.

Eiko I Fried. 2015. Problematic assumptions have slowed down depression research: why symptoms, not syndromes are the way forward. *Frontiers in psychology*, 6:309.

Chris D Frith. 2021. The neural basis of consciousness. *Psychological medicine*, 51(4):550–562.

Igor Galynker, Zimri S Yaseen, Abigail Cohen, Ori Benhamou, Mariah Hawes, and Jessica Briggs. 2017. Prediction of suicidal behavior in high risk psychiatric patients using an assessment of acute suicidal state: the suicide crisis inventory. *Depression and anxiety*, 34(2):147–158.

Nicole Geach, Nima G Moghaddam, and Danielle De Boos. 2018. A systematic review of team formulation in clinical psychology practice: definition, implementation, and outcomes. *Psychology and Psychotherapy: Theory, Research and Practice*, 91(2):186–215.

Michael SA Graziano. 2021. Understanding consciousness. *Brain*, 144(5):1281–1283.

E Clare Harris and Brian Barraclough. 1997. Suicide as an outcome for mental disorders. a meta-analysis. *British journal of psychiatry*, 170(3):205–228.

Simon Hatcher, Allison Crawford, and Nicole Coupe. 2017. Preventing suicide in indigenous communities. *Current opinion in psychiatry*, 30(1):21–25.

Richard Heeks. 2006. Health information systems: Failure, success and improvisation. *International journal of medical informatics*, 75(2):125–137.

M. A. Hom, M. C. Podlogar, I. H. Stanley, and T. E. Joiner. 2017. Ethical issues and practical challenges in suicide research. *Crisis*, 38(2):107–114.

R. F. Hunter, A. Gough, N. O'Kane, G. McKeown, A. Fitzpatrick, T. Walker, M. McKinley, M. Lee, and F. Kee. 2018. Ethical issues in social media research for public health. *Am J Public Health*, 108(3):343–348.

Lucy Johnstone. 2018. Psychological formulation as an alternative to psychiatric diagnosis. *Journal of Humanistic Psychology*, 58(1):30–46.

Lucy Johnstone and Rudi Dallos. 2013. *Introduction to formulation*, pages 21–37. Routledge.

Kenneth S Kendler. 2009. An historical framework for psychiatric nosology. *Psychological medicine*, 39(12):1935–1941.

R. C. Kessler, R. M. Bossarte, A. Luedtke, A. M. Zaslavsky, and J. R. Zubizarreta. 2020. Suicide prediction models: a critical review of recent research with recommendations for the way forward. *Mol Psychiatry*, 25(1):168–179.

Lina Keutzer and Ulrika SH Simonsson. 2020. Medical device apps: an introduction to regulatory affairs for developers. *JMIR mHealth and uHealth*, 8(6):e17567.

Cynthia F Kurtz and David J Snowden. 2003. The new dynamics of strategy: Sense-making in a complex and complicated world. *IBM systems journal*, 42(3):462–483.

Martin J La Roche, Milton A Fuentes, and Devon Hinton. 2015. A cultural examination of the dsm-5: Research and clinical implications for cultural minorities. *Professional Psychology: Research and Practice*, 46(3):183.

M. M. Large. 2018. The role of prediction in suicide prevention. *Dialogues Clin Neurosci*, 20(3):197–205.

Matthew Michael Large, Daniel Thomas Chung, Michael Davidson, Mark Weiser, and Christopher James Ryan. 2017. In-patient suicide: selection of people at risk, failure of protection and the possibility of causation. *BJPsych open*, 3(3):102–105.

Keri Rose Lawson-Te Aho Dr. 2017. The case for re-framing māori suicide prevention research in aotearoa/new zealand: Applying lessons from indigenous suicide prevention research. *Journal of Indigenous Research*, 6(2017):1.

Richard T Liu, Alexandra H Bettis, and Taylor A Burke. 2020. Characterizing the phenomenology of passive suicidal ideation: a systematic review and meta-analysis of its prevalence, psychiatric comorbidity, correlates, and comparisons with active suicidal ideation. *Psychological medicine*, 50(3):367–383.

Chris Mace and Sharon Binyon. 2005. Teaching psychodynamic formulation to psychiatric trainees: Part 1: Basics of formulation. *Advances in Psychiatric Treatment*, 11(6):416–423.

Narayana Manjunatha. 2019. Case presentation in academic psychiatry: The clinical applications, purposes, and structure of formulation and summary. *Indian Journal of Psychiatry*, 61(6):644.

Ian Marsh. 2018. *Historical phenomenology: Understanding experiences of suicide and suicidality across time*, pages 1–12. Springer.

Jeffery Martin, Jessica M LaCroix, Laura A Novak, and Marjan Ghahramanlou-Holloway. 2020. Typologies of suicide: A critical literature review. *Archives of suicide research*, 24(sup1):25–40.

Kahu McClintock and Rachel McClintock. 2017. Hoea te waka: Indigenous suicide prevention outcomes framework and evaluation processes-part 1. *Journal of Indigenous Wellbeing*, 2(2):68–76.

Catherine M McHugh, Amy Corderoy, Christopher James Ryan, Ian B Hickie, and Matthew Michael Large. 2019. Association between suicidal ideation and suicide: meta-analyses of odds ratios, sensitivity, specificity and positive predictive value. *BJPsych open*, 5(2).

Lindsey C McKernan, Ellen W Clayton, and Colin G Walsh. 2018. Protecting life while preserving liberty: ethical recommendations for suicide prevention with artificial intelligence. *Frontiers in psychiatry*, 9:650.

Giorgia Michelini, Isabella M Palumbo, Colin G DeYoung, Robert D Latzman, and Roman Kotov. 2021. Linking rdoc and hitop: A new interface for advancing psychiatric nosology and neuroscience. *Clinical psychology review*, 86:102025.

NEAC. 2019. National ethical standards for health and disability research and quality improvement. *Wellington: Ministry of Health*.

Barnaby Nelson, Patrick D McGorry, and Anthony V Fernandez. 2021. Integrating clinical staging and phenomenological psychopathology to add depth, nuance, and utility to clinical phenotyping: a heuristic challenge. *The Lancet Psychiatry*, 8(2):162–168.

Matthew K Nock, Franchesca Ramirez, and Osiris Rankin. 2019. Advancing our understanding of the who, when, and why of suicide risk. *JAMA psychiatry*, 76(1):11–12.

Martin Orr and Shankar Sankaran. 2007. Mutual empathy, ambiguity, and the implementation of electronic knowledge management within the complex health system. *Emergence: Complexity & Organization*, 9.

Sherry Pagoto and Camille Nebeker. 2019. How scientists can take the lead in establishing ethical practices for social media research. *Journal of the American Medical Informatics Association*, 26(4):311–313.

A. R. Pisani, D. C. Murrie, and M. M. Silverman. 2016. Reformulating suicide risk formulation: From prediction to prevention. *Acad Psychiatry*, 40(4):623–9.

Alejandro Porras-Segovia, Maria M Perez-Rodriguez, Pilar López-Esteban, Philippe Courtet, Jorge López-Castromán, Jorge A Cervilla, Enrique Baca-García, et al. 2019. Contribution of sleep deprivation to suicidal behaviour: a systematic review. *Sleep medicine reviews*, 44:37–47.

W Nicholson Price, Sara Gerke, and I Glenn Cohen. 2019. Potential liability for physicians using artificial intelligence. *Jama*, 322(18):1765–1766.

S. Pridmore. 2015. Mental disorder and suicide: a faulty connection. *Aust N Z J Psychiatry*, 49(1):18–20.

Bahlol Rahimi, Hamed Nadri, Hadi Lotfnezhad Afshar, and Toomas Timpka. 2018. A systematic review of the technology acceptance model in health informatics. *Applied clinical informatics*, 9(03):604–634.

Diana Rangihuna, Mark Kopua, and David Tipene-Leach. 2018. Mahi a atua: A pathway forward for māori mental health. *New Zealand Medical Journal*, 131(1471):79–83.

Jaco Renken and Richard Heeks. 2019. Champions of is innovations. *Communications of the Association for Information Systems*, 44(1):38.

Philip Resnik, April Foreman, Michelle Kuchuk, Katherine Musacchio Schafer, and Beau Pinkham. 2020. Naturally occurring language as a source of evidence in suicide prevention. *Suicide and Life-Threatening Behavior*.

C. J. Ryan. 2015. Suicide explained! *Aust N Z J Psychiatry*, 49(1):83–4.

Han-Chin Shing, Suraj Nair, Ayah Zirikly, Meir Friedenberg, Hal Daumé III, and Philip Resnik. 2018. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, pages 25–36.

Han-Chin Shing, Philip Resnik, and Douglas W Oard. 2020. A prioritization model for suicidality risk assessment. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8124–8137.

Dan J Stein and Geoffrey M Reed. 2019. Global mental health and psychiatric nosology: Dsm-5, icd-11, and rdoc. *Brazilian Journal of Psychiatry*, 41(1):3–4.

Mark Sullivan, Michael F Walton, Elizabeth L Auchincloss, and Julie B Penzner. 2020. Teaching psychiatric formulation to residents and faculty. *Academic Psychiatry*, pages 1–4.

George Szmukler. 2014. Fifty years of mental health legislation: Paternalism, bound and unbound. *Psychiatry: Past, present, and prospect*, pages 133–153.

Hamed Taherdoost. 2018. A review of technology acceptance and adoption models and theories. *Procedia manufacturing*, 22:960–967.

Myrthe L Tielman, Mark A Neerincx, Claudia Pagliari, Albert Rizzo, and Willem-Paul Brinkman. 2019. Considering patient safety in autonomous e-mental health systems–detecting risk situations and referring patients back to human care. *BMC medical informatics and decision making*, 19(1):47.

Michelle Torok, Jin Han, Simon Baker, Aliza Werner-Seidler, Iana Wong, Mark E Larsen, and Helen Christensen. 2020. Suicide prevention using self-guided digital interventions: a systematic review and meta-analysis of randomised controlled trials. *The Lancet Digital Health*, 2(1):e25–e36.

Leanne Townsend and Claire Wallace. 2016. Social media research: A guide to ethics. *University of Aberdeen*, pages 1–16.

Andrew S Tubbs, Michael L Perlis, and Michael A Grandner. 2019. Surviving the long night: the potential of sleep health for suicide prevention. *Sleep medicine reviews*, 44:83.

Kathryn Turner, Nicolas JC Stapelberg, Jerneja Sveticic, and Anthony R Pisani. 2021. Suicide risk classifications do not identify those at risk: where to from here? *Australasian psychiatry*, page 10398562211032233.

Clive Unsworth. 1991. Mental disorder and the tutelary relationship: from pre-to post-carceral legal order. *Journal of Law and Society*, 18(2):254–278.

Monika A Waszczuk, Mark Zimmerman, Camilo Ruggero, Kaiqiao Li, Annmarie MacNamara, Anna Weinberg, Greg Hajcak, David Watson, and Roman Kotov. 2017. What do clinicians treat: Diagnoses or symptoms? the incremental validity of a symptom-based, dimensional characterization of emotional disorders in predicting medication prescription patterns. *Comprehensive Psychiatry*, 79:80–88.

Priyanthy Weerasekera. 1993. Formulation: A multiperspective model. *The Canadian Journal of Psychiatry*, 38(5):351–358.

Matthew L Williams, Pete Burnap, Luke Sloan, Curtis Jessop, and Hayley Lepps. 2017. *Users' views of ethics in social media research: Informed consent, anonymity, and harm.* Emerald Publishing Limited.

R. Woodford, M. J. Spittal, A. Milner, K. McGill, N. Kapur, J. Pirkis, A. Mitchell, and G. Carter. 2019. Accuracy of clinician predictions of future self-harm: A systematic review and meta-analysis of predictive studies. *Suicide Life Threat Behav*, 49(1):23–40.

World Health Organization. 2014. *Preventing suicide: A global imperative.* World Health Organization.

Marianne Wyder, Manaan Kar Ray, Samara Russell, Kieran Kinsella, David Crompton, and Jeremy van den Akker. 2021. Suicide risk assessment in a large public mental health service: do suicide risk classifications identify those at risk? *Australasian Psychiatry*, 29(3):322–325.

Zimri S Yaseen, Mariah Hawes, Shira Barzilay, and Igor Galynker. 2019. Predictive validity of proposed diagnostic criteria for the suicide crisis syndrome: an acute presuicidal state. *Suicide and Life-Threatening Behavior*, 49(4):1124–1135.

Peter K Yu. 2020. The algorithmic divide and equality in the age of artificial intelligence. *Florida Law Review*, 72:19–44.

Mark Zimmerman, William Ellison, Diane Young, Iwona Chelminski, and Kristy Dalrymple. 2015. How many different ways do patients meet the diagnostic criteria for major depressive disorder? *Comprehensive psychiatry*, 56:29–34.

Ayah Zirikly, Philip Resnik, Ozlem Uzuner, and Kristy Hollingshead. 2019. Clpsych 2019 shared task: Predicting the degree of suicide risk in reddit posts. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 24–33.