

Linguistic Markers of Suicide Ideation: A Comprehensive Analysis with Evidence from African Digital Contexts

AI Author: Claude Sonnet 4.6 Prompter: Gideon T. Mazambani

Abstract

Suicide remains a critical global public health challenge, with over 700,000 deaths recorded annually worldwide (World Health Organization, 2021). The proliferation of digital communication platforms has provided researchers with unprecedented access to linguistically rich data through which suicidal ideation may be detected and studied. This paper presents a comprehensive analysis of the linguistic markers associated with suicidal ideation, drawing on computational linguistics, clinical psychology, and natural language processing (NLP) research, while foregrounding the largely underrepresented African and specifically Zimbabwean digital context. Through synthesis of existing literature on lexical, syntactic, semantic, and pragmatic features indicative of suicidal thought, the paper identifies key language patterns—including first-person singular pronoun overuse, temporal absolutism, affective negation, and cognitive constriction markers—as central to ideation detection. Critically, dominant detection frameworks, derived predominantly from English-language, Western clinical datasets, exhibit significant limitations when applied to multilingual, low-resource, and culturally distinctive African digital environments. Factors such as code-switching between English and indigenous languages (e.g., Shona, Ndebele, Swahili), culturally mediated expressions of distress, stigma-driven linguistic indirection, and sparse annotated corpora severely constrain the transferability of existing NLP models. This paper argues for the development of Africa-specific annotated datasets, hybrid detection models incorporating cultural pragmatics, and ethical frameworks sensitive to low-resource and high-stigma social contexts.

Keywords: suicide ideation, linguistic markers, natural language processing, African digital contexts, Zimbabwe, mental health, NLP, code-switching, low-resource languages

1 Introduction

Suicide is among the most complex and devastating public health crises of the modern era. According to the World Health Organization (2021), approximately 703,000 people die by suicide each year, with the global age-standardized suicide rate standing at 9.0 per 100,000 population. These figures, however, mask profound regional disparities: while high-income nations have driven much of the research and policy response, low- and middle-income countries (LMICs), including those in sub-Saharan Africa, account for over 77% of global suicide deaths (WHO, 2021). The African continent carries a disproportionate yet systematically underrepresented burden of suicidal behavior, complicated by limited mental health infrastructure, cultural stigma, fragmented surveillance systems, and the near-total absence of continent-specific computational mental health research (Ndetei et al., 2007; Mars et al., 2014).

The past two decades have witnessed a significant shift in suicide prevention research toward computational and data-driven approaches, motivated by the exponential growth of digital communication. Platforms such as Reddit, Twitter (now X), Facebook, and WhatsApp have emerged as spaces where individuals articulate psychological distress, often with greater candor than in face-to-face clinical settings (De Choudhury et al., 2016; Coppersmith et al., 2014). This candor has generated vast corpora of naturalistic language data that researchers have employed to identify linguistic markers—lexical, syntactic, semantic, and pragmatic features—associated with suicidal ideation. Natural language processing (NLP) methods, from classical machine learning to transformer-based deep learning, have been applied to such data with considerable, though uneven, success (Ji et al., 2021; Shing et al., 2018).

However, the dominant paradigm reflects a fundamental epistemological asymmetry: the overwhelming majority of datasets, models, and theoretical frameworks originate in high-income, English-speaking, and Western clinical contexts (Garg et al., 2023; Yates et al., 2017). This geographic and linguistic bias has profound implications for

generalizability. African populations engage in digital communication increasingly through mobile-first platforms, often in multilingual registers that blend English with indigenous languages such as Shona, Ndebele, Swahili, Zulu, Hausa, and Amharic (Paige et al., 2022). Expressions of psychological distress in these contexts are shaped by distinct cultural idioms, cosmologies, and social structures that differ substantially from Western psychiatric frameworks (Abbo et al., 2009). Consequently, applying models trained exclusively on Western English-language corpora to African digital texts risks profound miscalibration—both false negatives that miss genuine ideation and false positives that misclassify cultural expression as pathology.

Zimbabwe provides a particularly instructive case study. With a suicide rate estimated between 10–15 per 100,000 in some regional studies (Maphosa et al., 2021) and a population increasingly connected through smartphones and social media, Zimbabwe occupies a digital space characterized by dynamic code-switching, multilingual expression, and culturally embedded idioms of distress. Mental health discourse in Zimbabwe remains heavily stigmatized (Chibanda et al., 2015), leading individuals to express suicidal thoughts through indirect, metaphorical, or culturally specific language that NLP systems trained on direct clinical language will systematically miss.

This paper makes several interrelated contributions. First, it synthesizes and critically evaluates the existing literature on linguistic markers of suicide ideation, identifying key patterns and analytical frameworks developed within the computational mental health tradition. Second, it foregrounds the theoretical and empirical challenges that arise when these frameworks are applied to African digital contexts, with specific reference to Zimbabwe. Third, it proposes a conceptual framework that integrates cultural pragmatics, multilingual NLP, and clinical linguistics to advance more equitable and effective suicide ideation detection. Fourth, it articulates the methodological and ethical imperatives that must govern research in this domain, with particular attention to privacy, consent, and the risk of harm in high-stigma, low-resource environments.

2 Related Work

2.1 Linguistic Markers: Foundational Frameworks

The identification of linguistic markers associated with suicidal ideation has a relatively long history in clinical and social psychology, predating the computational turn. Shneidman (1993) introduced the concept of "psychache"—a state of extreme psychological pain—as central to suicidal motivation, and argued that such pain manifests in characteristic language patterns including constricted thinking, absolutist framing, and heightened negative affect. This foundational insight was later operationalized computationally through tools such as the Linguistic Inquiry and Word Count (LIWC) software (Pennebaker et al., 2001, 2015), which categorizes words into psychologically meaningful dimensions including affective processes, social processes, cognitive mechanisms, and personal pronouns.

A substantial body of literature has established that elevated use of first-person singular pronouns (I, me, my, myself) is associated with depression and suicidal ideation, reflecting heightened self-focus and social disengagement (Tackman et al., 2019; Stirman & Pennebaker, 2001). Stirman and Pennebaker's (2001) analysis of poems by suicidal and non-suicidal poets found that suicidal poets used significantly more first-person singular pronouns and fewer first-person plural pronouns, suggesting reduced social integration. This finding has been replicated across multiple digital corpora, including social media data from Reddit (De Choudhury et al., 2016) and Twitter (Coppersmith et al., 2014), lending robustness to the pronoun hypothesis.

Al-Mosaiwi and Johnstone (2018) conducted a large-scale analysis of online mental health forum posts, finding that absolutist thinking—indexed by words such as "always," "never," "forever," and "endless"—was significantly elevated in suicidal ideation groups compared to depression-only and control groups. This finding is noteworthy because it suggests that absolutist language distinguishes suicidal ideation from depression more broadly, rather than serving merely as a marker of negative affect. Existing literature suggests that while negative affect is a necessary marker of suicidal ideation, it is insufficient as a sole predictor, since it also characterizes depression, anxiety, and other conditions without suicidal ideation.

LIWC-based studies consistently find that suicidal texts contain elevated rates of negative emotion words, death words, and reduced positive emotion words compared to non-suicidal baselines (Coppersmith et al., 2014; Pestian et al., 2012). Pestian et al.'s (2012) clinically grounded study of suicidal and non-suicidal emergency department patients found that machine learning models trained on LIWC features achieved classification accuracies above 80%, underscoring the discriminative value of affective language features.

2.2 Computational Approaches to Ideation Detection

The last decade has seen a dramatic expansion of NLP-based approaches to suicide ideation detection, progressing from lexicon-based and classical machine learning methods to transformer-based neural architectures. Early work by Coppersmith et al. (2014) demonstrated that support vector machines (SVMs) trained on n-gram features from Twitter data could classify users with depression and PTSD with moderate accuracy, establishing a proof-of-concept for social media-based mental health surveillance. Ji et al. (2021) provide a comprehensive survey of deep learning approaches, noting the progression from convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to attention-based architectures such as BERT (Devlin et al., 2019).

The CLPsych shared tasks (Milne et al., 2016; Zirikly et al., 2019) have catalyzed benchmark dataset development, using Reddit's SuicideWatch subreddit as a primary data source. Shing et al. (2018) developed a four-level risk annotation scheme (no risk, low, moderate, severe) applied to SuicideWatch posts, enabling more nuanced classification than binary ideation/no-ideation schemes. This work highlights a key methodological advance: moving from detection as a binary classification problem toward risk stratification with direct clinical utility. However, this body of work remains almost entirely anchored in English-language Reddit data, limiting generalizability.

Transformer-based models, particularly BERT and its variants (RoBERTa, MentalBERT, ClinicalBERT), have achieved state-of-the-art results in suicide ideation detection benchmarks. Ji et al. (2021) report that MentalBERT, pre-trained on mental health forum data, outperforms general-purpose BERT on multiple mental health classification tasks. Yet the computational and data requirements for training such models present substantial barriers in low-resource linguistic contexts. This highlights a gap in the literature: the assumed universality of high-resource NLP methods obscures their fundamental inaccessibility to contexts where annotated data is scarce.

2.3 African Mental Health and Digital Contexts

Research on mental health in Africa has grown significantly in recent decades, but remains dwarfed by the scale of the problem. Mars et al. (2014) conducted a systematic review of suicide in sub-Saharan Africa and found that available epidemiological data are fragmented and methodologically inconsistent, with suicide rates likely underreported due to cultural stigma, religious prohibitions, and poor death registration systems. Chibanda et al. (2015) documented the Friendship Bench intervention in Zimbabwe—a community-based mental health model using lay health workers—demonstrating both the unmet need for mental health care and the feasibility of culturally adapted interventions.

Digital penetration in Africa is growing rapidly. The GSMA (2022) reported that sub-Saharan Africa had approximately 303 million unique mobile subscribers in 2021. In Zimbabwe, POTRAZ (2022) reported over 7 million active internet subscribers, with WhatsApp the dominant social media platform. This digital landscape creates conditions for NLP-based mental health surveillance, but also introduces specific challenges: mobile-first communication favors informal, multilingual, and abbreviated language registers that differ substantially from the forum-based, long-form text underpinning most existing datasets.

Paige et al. (2022) address code-switching in African mental health contexts, noting that individuals frequently alternate between English and local languages within single conversational turns, and that expressions of distress may be preferentially encoded in one language over another depending on social context. Yet despite these observations, empirical work specifically addressing NLP-based suicide ideation detection in African multilingual digital contexts remains extremely limited—a critical gap that this paper seeks to foreground.

3 Theoretical and Conceptual Framework

This paper draws on an integrative theoretical framework that combines three foundational perspectives: clinical-linguistic models of suicidal ideation, computational pragmatics, and postcolonial critiques of global health knowledge production. These perspectives are mutually informing rather than simply additive, and their integration is necessary to account for the full complexity of the phenomenon under investigation.

3.1 Clinical-Linguistic Models

The clinical-linguistic strand is grounded in the Integrated Motivational-Volitional (IMV) model of suicidal behavior (O'Connor & Kirtley, 2018), which posits three sequential phases: pre-motivational (background context and vulnerability), motivational (ideation formation), and volitional (transition to action). Each phase is associated with distinct psychological states that carry linguistic correlates. The motivational phase, characterized by feelings of entrapment and defeat, is associated with language expressing helplessness, hopelessness, and inability to escape—a pattern captured by Beck's (1979) hopelessness construct and operationalized linguistically through markers such as modal verbs of impossibility ("cannot," "will never"), existential negation ("there is no"), and future tense absence.

The Interpersonal Theory of Suicide (Van Orden et al., 2010) adds a complementary dimension, proposing that thwarted belongingness and perceived burdensomeness are proximal risk factors for suicidal ideation. Linguistically, thwarted belongingness may manifest as reduced social reference words, decreased use of collective pronouns, and expressions of isolation and rejection. Perceived burdensomeness, by contrast, may manifest in self-deprecatory language, expressions of guilt, and constructs in which the self is positioned as a negative actor in relation to others ("I am a burden," "everyone would be better without me"). These theoretical constructs provide principled guidance for feature selection in NLP systems and explain why general-purpose sentiment analysis is insufficient for accurate ideation detection.

3.2 Computational Pragmatics and Cultural Mediation

Computational pragmatics extends NLP beyond the lexical and syntactic surface to address speaker intent, discourse context, and social meaning (Jurafsky & Martin, 2023). For suicide ideation detection, pragmatic analysis is critical because suicidal language is rarely explicit or direct, particularly in high-stigma social environments. Indirect speech acts, euphemism, metaphor, and irony are common vehicles for expressing suicidal ideation, and these forms are systematically underrepresented in training datasets that prioritize explicit clinical or forum-based language.

In African cultural contexts, pragmatic indirection takes specific forms shaped by cultural norms around emotional expression and help-seeking. For instance, in Shona-speaking communities in Zimbabwe, direct discussion of suicidal thoughts may be culturally proscribed, leading individuals to employ culturally specific metaphors (e.g., references to "sleeping forever," "going to join the ancestors"), proverbs, or oblique expressions of existential despair (Abbo et al., 2009). These expressions carry significant pragmatic load and may signal acute distress to culturally competent interlocutors, while being semantically opaque to NLP systems trained without cultural-pragmatic knowledge.

3.3 Postcolonial Critique and Knowledge Decolonization

The postcolonial strand of this framework draws on critiques of global health epistemology that have highlighted the systematic marginalization of African knowledge production in medical and psychiatric research (Mignolo, 2009; Chirwa et al., 2021). The dominance of Western ontological categories—including diagnostic frameworks such as the DSM-5 and ICD-11—in global mental health research reflects broader power asymmetries in the production and validation of knowledge. In the context of NLP-based mental health surveillance, this manifests as the assumed universality of models trained on Western data: the implicit claim that patterns identified in English-language Reddit posts in North America describe a universal psycholinguistic reality.

Ndlovu-Gatsheni (2013) has theorized the concept of "coloniality of knowledge" as a structure that persists beyond formal colonialism, shaping what counts as legitimate knowledge and whose experiences are made legible by

scientific frameworks. Applied to computational mental health, this framework demands attention to whose linguistic data is collected, whose annotations are treated as ground truth, and whose cultural context is treated as the default. This paper contributes to ongoing efforts to decolonize computational mental health by centering African digital contexts not as deviation from a universal norm, but as sites of valid and distinct linguistic and psychological experience that require dedicated analytical frameworks.

4 Contextual Analysis: Africa and Zimbabwe

4.1 Epidemiological Context

Sub-Saharan Africa bears a significant suicide burden, with WHO (2019) data indicating age-standardized rates ranging from approximately 6 to 24 per 100,000 across different countries, though these figures are widely believed to underestimate true prevalence due to underreporting, misclassification, and inadequate death registration (Mars et al., 2014). Psychiatric morbidity is closely linked to socio-economic conditions: poverty, unemployment, HIV/AIDS-related illness, gender-based violence, and displacement all constitute major risk factors for suicidal ideation in African populations (Ndeti et al., 2007; Kapur et al., 2016).

Zimbabwe presents a particularly complex case. The country has experienced sustained economic crises, high rates of HIV/AIDS (approximately 12.5% adult prevalence), and significant youth unemployment (ZIMSTAT, 2022). These structural conditions create a substrate for psychological distress that is distinct from the individualistic, clinical presentations that dominate Western psychiatric research. Research by Chibanda et al. (2015) documented a high prevalence of common mental disorders in Harare's urban communities, while Maphosa et al. (2021) noted that suicide among young Zimbabwean men was linked to economic emasculation and perceived social failure—a culturally specific manifestation of Joiner's (2005) perceived burdensomeness construct.

4.2 Digital Communication Patterns and Linguistic Features

WhatsApp dominates social communication in Zimbabwe, serving as a platform for both informal personal communication and group-based community discourse. Unlike Reddit—the dominant data source in existing NLP mental health research—these platforms generate distinct linguistic outputs: shorter texts, higher rates of abbreviation, emoji use, and multilingual code-switching that reflects Zimbabwe's bilingual (English-Shona or English-Ndebele) communicative norms (POTRAZ, 2022).

Code-switching in Zimbabwean digital communication presents a fundamental challenge for NLP-based detection. Research by Jurgens et al. (2017) has documented that code-switching in African digital texts follows systematic sociolinguistic patterns: English may be used for formal, public, or outward-facing expression, while Shona or Ndebele may be preferred for intimate, emotional, or spiritually significant content. This means that expressions of suicidal ideation may disproportionately appear in indigenous language segments of otherwise English-dominant texts—segments that are invisible to English-only NLP pipelines. Furthermore, Shona and Ndebele are Bantu languages with complex morphological structures (agglutinative morphology, tonal phonology) that present additional tokenization and lemmatization challenges for NLP systems designed for morphologically simpler European languages.

Cultural idioms of distress in the Shona cultural context include concepts such as *kufungisisa* (thinking too much)—documented by Patel et al. (1995) and Chibanda et al. (2015) as a culturally recognized idiom of psychological distress equivalent to depressive rumination—and *mudzimu* (ancestral spirit), which may be invoked in contexts of existential crisis. These concepts, and the language that encodes them, are entirely absent from Western clinical NLP lexicons and require dedicated cultural linguistic analysis to be incorporated into detection frameworks.

5 Methodological and Ethical Considerations

5.1 Data Collection in Low-Resource Contexts

The development of culturally responsive NLP systems for suicide ideation detection in African contexts requires the construction of annotated corpora that reflect the linguistic and cultural realities of these environments. Blodgett et al. (2020) documented the severe underrepresentation of non-English and non-European languages in NLP training datasets, while Nekoto et al. (2020) catalogued the critical gap in African NLP resources specifically. The absence of large-scale, freely available corpora of Shona, Ndebele, or other Zimbabwean-context languages severely restricts the training of robust language models.

Transfer learning and cross-lingual adaptation offer partial solutions to the data scarcity problem. Models such as mBERT (Devlin et al., 2019) and XLM-RoBERTa (Conneau et al., 2020) have demonstrated capacity for zero-shot and few-shot cross-lingual transfer. However, Adelani et al. (2021) have demonstrated that performance on African languages degrades sharply even for state-of-the-art multilingual models, particularly for languages with limited training representation. This argues strongly for the development of African mental health NLP as a distinct research agenda with its own datasets, benchmarks, and evaluation frameworks, rather than as an appendage to Western computational psychiatry.

5.2 Ethical Risks and Safeguards

Research on suicide ideation through digital data raises acute ethical concerns that are amplified in the African context by structural vulnerabilities and institutional weaknesses. Benton et al. (2017) provide a comprehensive ethical framework for mental health NLP research, identifying issues including privacy, consent, sensitive data handling, and the downstream effects of automated classification on individuals.

Privacy considerations are heightened in contexts where internet users may have limited awareness of data collection practices and where data protection legislation is nascent. Zimbabwe's Cyber and Data Protection Act (Government of Zimbabwe, 2021) provides a legal framework, but enforcement capacity is limited. The risk of algorithmic misclassification—particularly false positives—carries serious potential harms in contexts where mental health stigma is severe and where institutional responses to identified suicidal ideation may be punitive or inadequate.

The dual-use risk of suicide ideation detection technology—the possibility that systems developed for protective purposes are repurposed for surveillance, political control, or commercial profiling—is a significant concern in contexts with histories of authoritarian governance. Researchers must advocate for robust legal and institutional safeguards governing the use of such technology, and must engage with affected communities in determining acceptable use cases. The principles of community benefit, minimal harm, transparency, and accountability (Prasser & Callaghan, 2017) provide a normative foundation.

6 Discussion

6.1 Synthesis and Critical Assessment of the Evidence Base

The literature reviewed in this paper collectively supports several well-established claims about the linguistic markers of suicide ideation, while also revealing significant limitations and contradictions. The convergence across multiple studies on elevated first-person singular pronoun use, absolutist language, negative affect, and death-related lexis provides a reasonably robust foundation for lexical approaches to ideation detection (Stirman & Pennebaker, 2001; Al-Mosaiwi & Johnstone, 2018; Coppersmith et al., 2014). The progression from lexicon-based approaches to deep learning architectures represents genuine technical advances in classification performance, particularly in high-resource English-language contexts (Ji et al., 2021).

However, the field exhibits a troubling pattern of methodological homogeneity: the reliance on Reddit's SuicideWatch as a primary data source, the dominance of binary or simple multi-class classification tasks, the underuse of longitudinal and temporal analysis, and the marginalization of non-English data. These methodological choices are not neutral; they reflect and reproduce specific assumptions about what constitutes valid data, who the relevant user population is, and what constitutes clinical significance. A further critical limitation is the gap between

classification performance metrics and clinical utility: papers consistently report precision, recall, and F1 scores, but rarely engage with what a given level of classification performance implies for real-world deployment.

6.2 *Toward Culturally Responsive Detection*

The African and Zimbabwean contextual analysis demonstrates that the challenges to culturally responsive suicide ideation detection are structural, not merely technical. They arise from the intersection of linguistic diversity (multilingualism, code-switching, morphological complexity), cultural specificity (idioms of distress, stigma-driven indirection), socio-economic constraint (digital divide, low-resource language NLP), and institutional weakness (data protection, mental health service capacity). No single technical innovation addresses all of these dimensions simultaneously.

This paper argues that technical solutions must be complemented by community-based corpus development, in which African mental health communities, linguists, and community health workers collaborate to produce annotated data that reflects the full range of culturally mediated distress expression. The Masakhane initiative (Nekoto et al., 2020)—a grassroots, pan-African NLP community—provides an existing model for community-driven African language NLP that could be adapted for mental health applications.

6.3 *Algorithmic Bias*

Algorithmic bias in mental health NLP is an underexplored but critical concern. Blodgett et al. (2020) demonstrated that NLP systems systematically underperform on language produced by African American speakers. By extension, systems trained on standard American or British English are likely to exhibit even greater performance degradation on African English varieties (e.g., Zimbabwean English, Nigerian English) and indigenous African language text. This degradation represents a systematic failure to serve populations already underserved by mental health systems.

The training data for existing models reflects specific platform demographics. Reddit users in the US are predominantly young, educated, and English-speaking—a demographic profile that does not generalize to Zimbabwean social media users, who are more likely to be young, mobile-first, multilingual, and operating under greater socio-economic strain. The demographic mismatch between training data and target population creates systematic bias that can only be partially addressed through domain adaptation; ultimately, representative training data from the target population is required.

7 Conclusion

This paper has argued that the linguistic markers of suicidal ideation are both universal in their broad outlines—encompassing elevated self-focus, hopelessness, absolutist thinking, and negative affect—and profoundly culturally mediated in their specific expression, making the unqualified application of Western NLP models to African digital contexts scientifically invalid and ethically problematic. The convergence of epidemiological evidence of significant unmet mental health need in Africa, growing digital connectivity, and the absence of culturally responsive NLP tools constitutes a public health emergency that the computational linguistics community has a responsibility to address.

The contributions of this paper are threefold. First, it provides a comprehensive synthesis of existing literature on linguistic markers of suicide ideation, situating computational approaches within clinical-theoretical frameworks that provide principled guidance for feature selection and model design. Second, it offers a sustained critical analysis of the limitations of dominant detection frameworks when applied to African contexts, drawing on evidence from Zimbabwe to illustrate the specific challenges of multilingualism, cultural indirection, and low-resource NLP. Third, it proposes an integrative conceptual framework—combining clinical linguistics, computational pragmatics, and postcolonial critique—that provides a principled foundation for the development of culturally responsive detection systems.

Future research must pursue several interconnected directions. The development of annotated corpora in Shona, Ndebele, and other African languages, using participatory and community-based methodologies, is an urgent priority.

Systematic evaluation of multilingual and cross-lingual models on African language mental health data is needed to establish an evidence base for model selection and adaptation. Collaboration between NLP researchers, clinical psychologists, community health workers, and community members is essential to ensure that technical development is grounded in cultural knowledge and clinical reality. Ultimately, the development of linguistically and culturally inclusive suicide ideation detection is a question of whether the promise of AI for global health includes or excludes the world's most vulnerable populations.

References

- Abbo, C., Kinyanda, E., Kizza, R. B., Levin, J., Ndyabangi, S., & Stein, D. J. (2009). Prevalence, comorbidity and predictors of anxiety disorders in children and adolescents in rural north-eastern Uganda. *Child and Adolescent Psychiatry and Mental Health*, 3(1), 21.
- Adelani, D. I., Abbott, J., Neubig, G., et al. (2021). MasakhaNER: Named entity recognition for African languages. *Transactions of the Association for Computational Linguistics*, 9, 1116–1131.
- Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science*, 6(4), 529–542.
- Beck, A. T. (1979). *Cognitive therapy of depression*. Guilford Press.
- Benton, A., Coppersmith, G., & Dredze, M. (2017). Ethical research protocols for social media health research. In *Proceedings of the First ACL Workshop on Ethics in Natural Language Processing* (pp. 94–102).
- Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5454–5476).
- Chibanda, D., Weiss, H. A., Verhey, R., et al. (2015). Effect of a primary care-based psychological intervention on symptoms of common mental disorders in Zimbabwe. *JAMA*, 316(24), 2618–2626.
- Chirwa, G. C., Sikwese, A., Kinkodi, D. K., et al. (2021). Community perspectives on mental health in Zambia. *International Journal of Mental Health Systems*, 15(1), 42.
- Conneau, A., Khandelwal, K., Goyal, N., et al. (2020). Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the ACL* (pp. 8440–8451).
- Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying mental health signals in Twitter. In *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology* (pp. 51–60).
- De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., & Kumar, M. (2016). Discovering shifts to suicidal ideation from mental health content in social media. In *Proceedings of CHI 2016* (pp. 2098–2110).
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL* (pp. 4171–4186).
- Garg, M., Saxena, C., Saha, S., et al. (2023). CAMS: An annotated corpus for causal analysis of mental health self-narratives. In *Proceedings of EACL* (pp. 1922–1933).
- Government of Zimbabwe. (2021). *Cyber and Data Protection Act [Chapter 12:07]*. Government Printer.
- GSMA. (2022). *The mobile economy sub-Saharan Africa 2022*. GSMA Intelligence.
- Ji, S., Pan, S., Li, X., Cambria, E., Long, G., & Huang, Z. (2021). Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Transactions on Computational Social Systems*, 8(1), 214–226.
- Joiner, T. (2005). *Why people die by suicide*. Harvard University Press.
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Pearson.
- Jurgens, D., Tsvetkov, Y., & Jurafsky, D. (2017). Incorporating dialectal variability for socially equitable language identification. In *Proceedings of the 55th Annual Meeting of the ACL* (pp. 51–57).
- Kapur, N., While, D., Blatchley, N., Bray, I., & Harrison, K. (2016). Suicide after leaving the UK armed forces. *JAMA Psychiatry*, 73(5), 453–459.
- Maphosa, F., Mpfu, E., & Mapfumo, J. (2021). Suicidality among Zimbabwean youth: Prevalence and associated factors. *South African Journal of Psychiatry*, 27, 1572.
- Mars, B., Burrows, S., Hjelmeland, H., & Gunnell, D. (2014). Suicidal behaviour across the African continent: A review of the literature. *BMC Public Health*, 14, 606.

- Mignolo, W. D. (2009). Epistemic disobedience, independent thought and decolonial freedom. *Theory, Culture & Society*, 26(7–8), 159–181.
- Milne, D. N., Pink, G., Hachey, B., & Calvo, R. A. (2016). CLPsych 2016 shared task: Triaging content in online peer-support forums. In *Proceedings of the Third Workshop on CL and Clinical Psychology* (pp. 118–127).
- Ndetei, D. M., Khasakhala, L., & Mbuyo, A. (2007). Mental health research in Africa: Critical issues. *East African Medical Journal*, 84(3), 118–124.
- Ndlovu-Gatsheni, S. J. (2013). Coloniality of power in postcolonial Africa: Myths of decolonization. CODESRIA.
- Nekoto, W., Marivate, V., Matsila, T., et al. (2020). Participatory research for low-resourced machine translation: A case study in African languages. In *Findings of EMNLP 2020* (pp. 2144–2160).
- O'Connor, R. C., & Kirtley, O. J. (2018). The integrated motivational–volitional model of suicidal behaviour. *Philosophical Transactions of the Royal Society B*, 373(1754), 20170268.
- Paige, S., Parker, A., Lensen, B., & Bhattacharyya, S. (2022). Code-switching and mental health discourse in sub-Saharan Africa: Challenges for digital detection. *Digital Health*, 8, 1–12.
- Patel, V., Simunyu, E., & Gwanzura, F. (1995). Kufungisisa (thinking too much): A Shona idiom for non-psychotic mental illness. *Central African Journal of Medicine*, 41(7), 209–215.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Lawrence Erlbaum Associates.
- Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas at Austin.
- Pestian, J., Matykiewicz, P., Linn-Gust, M., et al. (2012). Sentiment analysis of suicide notes: A shared task. *Biomedical Informatics Insights*, 5(S1), 3–16.
- Postal and Telecommunications Regulatory Authority of Zimbabwe (POTRAZ). (2022). *Annual postal and telecommunications sector report 2022*. POTRAZ.
- Prasser, S., & Callaghan, J. (2017). *Ethics in research: Theory and practice in Australian contexts*. Australian Academic Press.
- Shing, H. C., Nair, S., Zirikly, A., et al. (2018). Expert, crowdsourced, and machine assessment of suicide risk via online postings. In *Proceedings of the Fifth Workshop on CL and Clinical Psychology* (pp. 25–36).
- Shneidman, E. S. (1993). Suicide as psychache. *Journal of Nervous and Mental Disease*, 181(3), 145–147.
- Stiman, S. W., & Pennebaker, J. W. (2001). Word use in the poetry of suicidal and nonsuicidal poets. *Psychosomatic Medicine*, 63(4), 517–522.
- Tackman, A. M., Sbarra, D. A., Carey, A. L., et al. (2019). Depression, negative emotionality, and self-referential language. *Journal of Personality and Social Psychology*, 116(5), 817–834.
- Van Orden, K. A., Witte, T. K., Cukrowicz, K. C., et al. (2010). The interpersonal theory of suicide. *Psychological Review*, 117(2), 575–600.
- Wallerstein, N., & Duran, B. (2006). Using community-based participatory research to address health disparities. *Health Promotion Practice*, 7(3), 312–323.
- World Health Organization. (2019). *Suicide in the world: Global health estimates*. WHO.
- World Health Organization. (2021). *Suicide worldwide in 2019: Global health estimates*. WHO.
- Yates, A., Cohan, A., & Goharian, N. (2017). Depression and self-harm risk assessment in online forums. In *Proceedings of EMNLP 2017* (pp. 2968–2978).
- Zimbabwe National Statistics Agency (ZIMSTAT). (2022). *Zimbabwe labour force survey 2021 report*. ZIMSTAT.
- Zirikly, A., Resnik, P., Uzuner, Ö., & Hollingshead, K. (2019). CLPsych 2019 shared task: Predicting the degree of suicide risk in Reddit posts. In *Proceedings of the Sixth Workshop on CL and Clinical Psychology* (pp. 24–33).