Mapping Digital Mental Health Discourse in Hong Kong's Online Forums: A Mixed-Methods Study of Stigma, Disclosure, and Peer Support for AI-Enabled Public Health Interventions

Celine Song

Background and Rationale

In high-pressure urban societies like Hong Kong, mental health stigma and cultural concerns around "face" (面子) often suppress open expression of psychological distress. Anonymous, user-driven online forums—such as LIHKG, Baby Kingdom, and DiscussHK—have become linguistically rich and socially meaningful spaces for self-disclosure, emotional expression, and peer interaction, particularly in the absence of culturally attuned formal mental health channels. Each of these forums serves a distinct generational and socio-cultural niche: LIHKG attracts digitally native youth engaged in anonymous, affectively charged discourse; Baby Kingdom provides a semi-private space where parenting-related mental load and emotional stress are frequently shared; and DiscussHK offers a broad-based platform where middle-aged and older adults participate in more moderated, deliberative conversations. Together, these platforms form a complementary ecology for understanding how mental health discourse circulates across age cohorts, social identities, and communicative styles in Hong Kong.

Research Design and Methodological Innovation

This study investigates how mental health discourse unfolds across these forums and explores how such discourse can inform the design of culturally sensitive, AI-enabled public health interventions. Employing a mixed-methods design, the study integrates computational social science methods with qualitative inquiry.

1. Computational and Quantitative Approaches

In the quantitative phase, this study employs a multi-layered computational framework to analyse mental health discourse at scale across three linguistically and culturally distinct online communities. At the core of the topic modelling pipeline is BERTopic, a transformer-based framework tailored with Cantonese-specific embeddings trained on forum-native data. This adaptation allows the model to capture culturally nuanced expressions of psychological distress—such as indirect complaint, affective self-deprecation, or metaphor-laden disclosures—that would be flattened or misclassified by conventional English-trained language models.

To analyse emotional tone, sentiment classification is conducted using large language models fine-tuned to detect subtle affective cues typical in Hong Kong's code-mixed online discourse. These include not only lexical sentiment (e.g., sadness, anger), but also discursive signals such as sarcasm, tonal ambivalence, and euphemistic expressions of distress that often evade rule-based or polarity-centric sentiment systems. The outputs are validated through a cross-lingual human annotation process, ensuring semantic alignment between AI-generated labels and culturally situated human judgments.

Beyond static measures of topic and sentiment prevalence, this study implements thread-level network analysis to uncover the temporal dynamics and relational contours of mental health talk. By modelling reply structures as discourse networks, we trace how stigma-related themes emerge, diffuse, and are challenged or reinforced through user interaction. Network centrality,

clustering coefficients, and cascade depth are used to characterize the embeddedness and resonance of emotional expressions within evolving conversations.

In parallel, linguistic feature extraction techniques are applied to model how stylistic and structural variations influence users' positioning in discussions. Features include lexical diversity (e.g., type-token ratio), syntactic complexity (e.g., clause density, dependency distance), and code-switching intensity (e.g., frequency and position of language alternation). These features are then linked to thematic roles (e.g., discloser, responder, empathizer) and affective profiles to identify communicative patterns associated with stigma resistance, support-seeking, and emotional disengagement. Together, these techniques offer a computationally rich, context-sensitive lens through which to map digital mental health discourse in Hong Kong.

2. Qualitative Inquiry and Narrative Analysis

The qualitative phase deepens and contextualizes the computational findings through discourse analysis of selected discussion threads. Rather than focusing on frequency or polarity, this analysis attends to the processual unfolding of conversations—examining how users frame psychological distress, position themselves and others in relation to stigma, and negotiate the boundaries between emotional vulnerability and social legitimacy.

Special attention is given to rhetorical strategies such as metaphor, irony, narrative pacing, and affective alignment, which serve as key mechanisms of emotional meaning-making and stigma navigation in multilingual, informally structured interactions. These discourse features—often fragmented, implicit, and culturally coded—reveal how users construct legitimacy, empathy, and solidarity over time.

To enhance both analytical efficiency and interpretive depth, large language models are used as interpretive aides. They assist in clustering semantically rich passages, identifying implicit stance markers, and generating summaries of emotionally complex exchanges. Their role is not to replace human judgment, but to extend its reach—surfacing latent narrative patterns and emotional dynamics that might otherwise remain unnoticed in large-scale data.

In complement to the computational phase, this inquiry captures the how of mental health discourse—how distress is narrated, how stigma is resisted, and how informal peer support is relationally co-constructed. It illuminates culturally specific modes of emotional expression and communicative care, revealing how online forums function not just as sites of expression, but as evolving vernacular infrastructures for resilience, recognition, and mental health sense-making.

Theoretical and Practical Contributions

This study offers both conceptual and applied contributions. Theoretically, it advances a culturally embedded, discourse-aware framework for analysing mental health communication in high-context, multilingual societies—addressing a major gap in stigma and disclosure research, which remains predominantly Western and monolingual. Methodologically, it refines the application of large language models by incorporating code-switch-aware embeddings, narrative-level semantic analysis, and human validation loops, enhancing both analytical granularity and cultural fidelity. Practically, the study demonstrates how anonymized, user-generated discourse can be ethically leveraged to detect early signs of emotional distress, identify communicative barriers to help-seeking, and model the dynamics of informal peer support at scale. Rather than positioning AI as a diagnostic authority, the study

proposes its role as a culturally adaptive augmentation to existing public health infrastructures—capable of generating empathetic, context-sensitive prompts, enhancing digital triage systems, and supporting stigma-reduction campaigns. Ultimately, this research contributes to building AI-enabled mental health interventions that are data-driven, ethically grounded, and responsive to the emotional realities of digitally mediated publics.