

Using natural language processing to detect mental health crises

January 17 2024, by Prabha Kannan

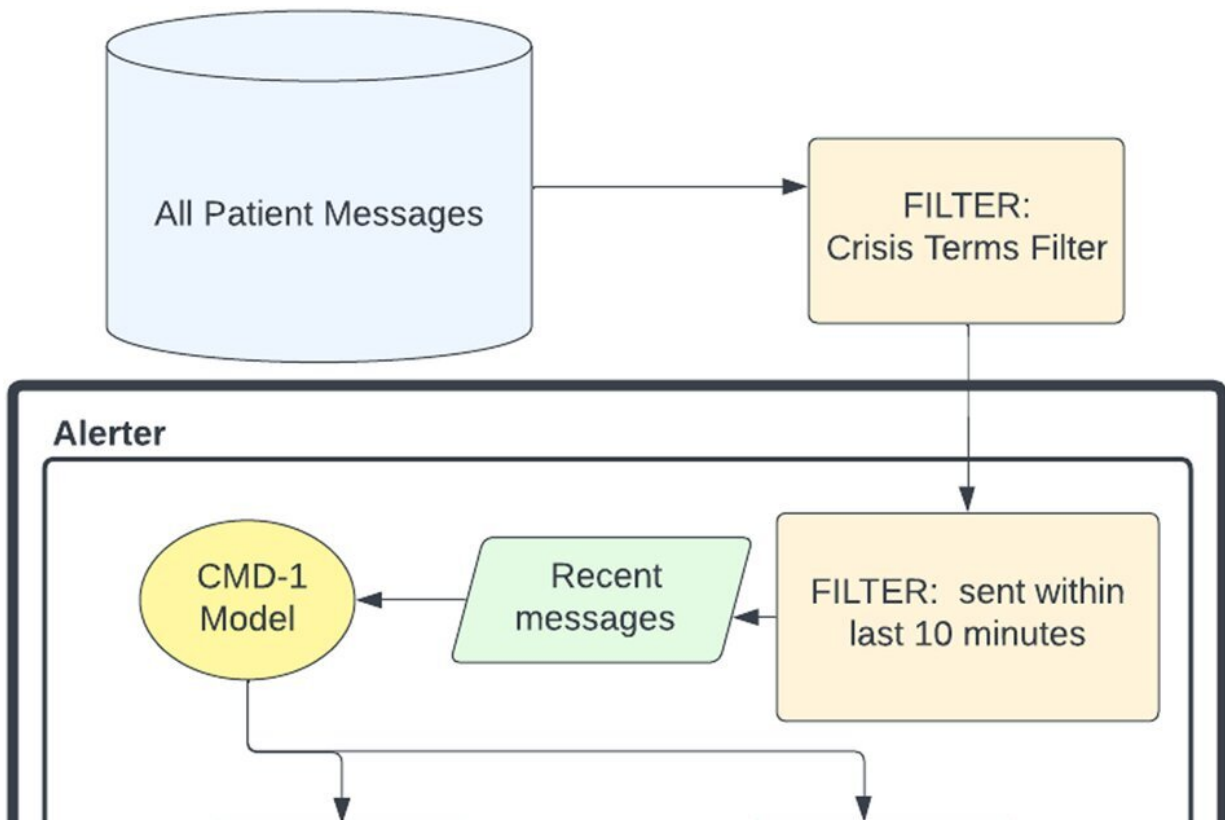


Diagram depicting the flow of patient messages through the CMD-1 system in model deployment. Credit: *npj Digital Medicine* (2023). DOI: 10.1038/s41746-023-00951-3. *npj Digital Medicine* (2023). DOI: 10.1038/s41746-023-00951-3

Mental health needs are on the rise. Today one in five Americans lives

with a mental health condition, and suicide rates over the last two decades have increased by more than 30%. Organizations like the National Alliance on Mental Illness (NAMI), which offers free support for those experiencing a crisis, saw a 60% increase in help-seekers between 2019 and 2021.

To address this increase, organizations and health care providers are turning to digital tools. Technology platforms like [crisis](#) hotlines, text lines, and online chat lines trained and added staff to support patients in crisis. Even so, dropped call rates for such organizations remain as high as 25%. Furthermore, the vast majority of these services are siloed from the callers' clinicians.

One of the primary factors contributing to this high drop rate is that patient demand significantly outnumbers available responders. In 2020, the National Suicide Prevention Lifeline reported a response rate of just 30% for chat and 56% for text messages, leaving many patients in crisis without support. Furthermore, these systems use a standard queuing method for incoming messages, where patients are served on a first-come-first-served basis, as opposed to their level of urgency.

What if these platforms could instead distinguish between urgent and non-urgent messages, thereby improving efficiency in triaging crisis cases?

This is precisely what Akshay Swaminathan and Ivan Lopez—Stanford [medical students](#)—set out to do with their team of interdisciplinary collaborators, including clinicians and operational leaders at Cerebral, a national online mental health company, where Swaminathan leads data science.

The research team includes Jonathan Chen, Stanford HAI affiliate and assistant professor of medicine in the Stanford Center for Biomedical

Informatics Research, and Olivier Gevaert, Stanford associate professor of medicine and of biomedical data science. The authors published their work in [*npj Digital Medicine*](#).

Using [natural language processing](#), the team developed a machine learning (ML) system called Crisis Message Detector 1 (CMD-1) that can identify and auto-triage concerning messages, reducing patient wait times from 10 hours to under 10 minutes.

"For clients who are suicidal, the wait time was simply too long. The implication of our research is that [data science](#) and ML can be successfully integrated into clinician workflows, leading to dramatic improvements when it comes to identification of patients at risk, and automating away these really manual tasks," Swaminathan says.

Their results underscore the importance of CMD-1's application to scenarios where speed is crucial. Lopez says, "CMD-1 enhances the efficiency of crisis response teams, allowing them to address a greater number of cases more effectively. With quicker triage, resources can be allocated more effectively, prioritizing urgent cases."

Empowering crisis specialists

The data that the team used as the foundation for CMD-1 came from Cerebral, which receives thousands of patient messages per day in its chat system. Messages can include such varied topics as appointment scheduling and medication refills, plus messages from patients in emergent crises.

Beginning with a random sample of 200,000 messages, they labeled patient messages as "crisis" or "non-crisis" using a filter that includes factors like key crisis words and patient IDs that had previously reported a crisis within the last week. Crisis messages warranting further attention

note expressions of suicidal or homicidal ideation, domestic violence, or non-suicidal self-injury (self-harm).

"For messages that are ambiguous, something like, 'I need help,' we erred on the side of calling it a crisis. From that phrase alone you don't know whether the patient needs help with scheduling an appointment or getting out of bed," Swaminathan says.

The team was also steadfast in ensuring that their approach would complement—but not replace—human review. CMD-1 surfaces crisis messages and sends them to a human to review as part of their typical crisis response workflow in a Slack interface. Any true crisis messages that the model fails to surface (false negatives) are reviewed by humans as part of the routine chat support workflow.

As Lopez says, "This approach is crucial in ensuring we reduce the risk of false negatives as much as possible. Ultimately, the human element in reviewing and interpreting messages ensures a balance between technological efficiency and compassionate care, which is essential in the context of mental health emergencies."

Given the sensitivity of the subject area, the team was extremely conservative with how messages were categorized. They considered False Negatives (missing a true crisis message) and False Positives (incorrectly surfacing a non-crisis message), and, working with the clinical stakeholders, determined that the cost of missing a False Negative was 20 times more undesirable than addressing a False Positive.

"This is a really key point when it comes to deploying ML models. Any ML model that's making a classification—calling it 'crisis' or 'not crisis'—first has to output a probability from zero to one. It outputs a probability that the message is a crisis, but we have to pick that

threshold, above which the model calls it a crisis and below which the model says it's not a crisis. The choice of that cutoff is a critical decision, and that decision should not be made by the people building the model, it should be made by the end users of the model. For us, that's the clinical teams," Swaminathan says.

Remarkably, CMD-1 was able to detect high-risk messages with impressive accuracy (97% sensitivity and 97% specificity), and the team reduced response time for help-seekers from over 10 hours to just 10 minutes. This speed is critical, as rapid intervention has the potential to redirect high-risk patients away from suicide attempts.

ML's potential in health care

Given their noteworthy results, the team is hopeful that more [machine learning](#) models will be deployed in health care settings, which is currently a rare occurrence, as model deployment requires careful translation to clinical settings and close attention to technical and operational considerations, as well as technological infrastructure.

"Often, data scientists create highly accurate ML models without fully addressing the pain points of stakeholders. As a result, these models, although technically proficient, may create additional work or fail to integrate seamlessly into existing clinical workflows. For more widespread adoption, data scientists must involve health care professionals from the outset, ensuring that the models address the challenges they're designed to solve, streamline rather than complicate tasks, and fit organically within the existing clinical infrastructure," says Lopez.

The development team for CMD-1 took a unique approach of assembling a cross-functional team of clinicians and data scientists to ensure the model met key clinical thresholds and could drive meaningful

outcomes in a real clinical operations setting. "The compelling system developed here demonstrates not just an analytical result, but the much harder work of integrating that into a real workflow that enables human patients and clinicians to reach each other at a critical moment," says Chen.

This cross-functional approach, combined with CMD-1's noteworthy results, showed Swaminathan and Lopez how technology could be used to augment the impact of clinicians. "This is the direction that AI in medicine is headed in, where we're using data to make health care delivery more human, to make clinicians' lives easier, and to empower them to deliver higher quality care," says Swaminathan.

More information: Akshay Swaminathan et al, Natural language processing system for rapid detection and intervention of mental health crisis chat messages, *npj Digital Medicine* (2023). [DOI: 10.1038/s41746-023-00951-3](https://doi.org/10.1038/s41746-023-00951-3)

Provided by Stanford University

Citation: Using natural language processing to detect mental health crises (2024, January 17) retrieved 27 April 2024 from <https://medicalxpress.com/news/2024-01-natural-language-mental-health-crises.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.