SERVE ADVISORY GROUP

WHITE PAPER | VSIGNALS PROJECT

FROM VSIGNALS TO E-SOLUTIONS: AN ANALYSIS OF VETERAN SUICIDE IDEATION ALERTS

ABSTRACT

In this research paper, we conduct an in-depth examination of respondent comments gathered via the Veteran's Crisis Line (VCL) VSignals program. The SERVE Advisory Group, operating as a key component of the RB Management Consultants team, offered contract support from January to September 2023.The study encompasses data collection, privacy protection, categorization utilizing Artificial Intelligence (AI): We deploy open-source AI tooling to produce a successful proof-of-concept trained to predict patient outcomes, while a pre-trained large language model (LLM) was used for zero-shot inference to assist in data analysis primarily utilizing Microsoft Excel. The primary aim is to categorize and extract meaningful insights from veteran's experiences, challenges, and concerns by classifying alerts into three major categories of Care, Services, and Benefits, and analyzing data to identify demographic trends. Subcategories were created, classified, and analyzed. Care was subcategorized into mental health care, physical care and treatment, or medication issues; Services was subcategorized into appointment issues, quality and trust in care, or the phone system; and Benefits included all survey responses that related to VA benefits or payments for care. Additionally, a category called "Other" was introduced to encompass the alerts that did not fit into an aforementioned category.

The study reveals that the data is complex and incomplete, making it difficult to extract important insights from all the veterans' survey responses. Additionally, the paper explores how Medallia, a company specializing in customer experience management software, can help improve the efficiency and accuracy of alert categorization and analysis, which can lead to better decision-making in crisis intervention.

BACKGROUND:

Since its founding in July 2007, the Veterans Crisis Line (VCL) has played a pivotal role in safeguarding the well-being of our nation's veterans. Over the years, this lifeline has been a beacon of hope, effectively managing an astounding 7.1 million calls. These calls have been a lifeline for approximately 287,000 veterans facing imminent crises, ensuring that vital emergency services are dispatched promptly, providing the support and care they urgently require. The Veterans Chat service, introduced in 2009, has been a digital haven for those seeking assistance and understanding. With unwavering dedication, it has addressed over 862,000 chat requests, demonstrating the importance of offering support through online channels. Recognizing the evolving needs and preferences of veterans, the Crisis Line expanded its reach into the world of texting in 2011. Since its inception, the Crisis Line texting service has diligently handled nearly 327,000 text requests, ensuring that no veteran's cry for help goes unanswered, whether through the spoken word or written text[1].

In fiscal year 2018, the Department of Veteran Affairs allocated $12.2 million to suicide prevention outreach, dedicating $1.5 million to a creative collaboration with Johnson & Johnson to launch the "No Veteran Left Behind" public service announcement featuring Tom Hanks on social media[2]. That year, during Suicide Prevention Month (September), these efforts garnered over 347,000 visits to the VCL website, demonstrating the program's substantial impact in engaging and educating the public[3].

In 2018, the Department of Veterans Affairs and the Veterans Crisis Line (VCL) introduced the VSignals program, aimed at assisting and supporting veterans nationwide. VCL addresses a wide spectrum of concerns through VSignals, including but not limited to suicidal thoughts, depression, anxiety, and self-harm issues. Veterans' experiences and needs relating to suicidal ideation, self-harm, and mental health concerns are communicated to the Veterans Crisis Line through trust surveys. Over 250,000 trust surveys are issued quarterly by the VA, as the result of an interaction with the Veteran so that the VA can collect feedback in various arenas such as appointment scheduling, quality of care, wait times, prescriptions, and much more. The trust surveys allow free-text responses and undergo keyword screening via a software program called Medallia; a software platform and company that specializes in customer experience management. The Medallia system is designed to help organizations collect, analyze, and act on customer feedback and data to improve their customer experiences and drive business success. It provides tools for gathering feedback through various channels, such as surveys, social media, and other customer interactions. The key components of the Medallia system typically include feedback collection, feedback analysis, reporting and dashboards, action management, and customer experience improvement. Medallia is commonly used in industries such as retail, hospitality, healthcare, and financial services to gain a better understanding of customer satisfaction and loyalty. It's a valuable tool for businesses looking to enhance customer relationships and drive better outcomes[4]. Utilizing Medallia for the VSignals program, respondent comments of surveys containing suicide-related keywords spark the creation of an alert, which is reviewed by the Data Entry Operator (DEO) for further assessment.

As of 2022, bipartisan backing for Veteran suicide prevention remains steady across different presidential administrations. The 2022 Annual VA budget request underscores this dedication, allocating $590 million exclusively for suicide prevention and outreach, primarily directed at augmenting VCL's resources and capabilities[5]. For FY 2022, $256 million was budgeted for VCL[6]. These initiatives harmonize with the overarching 10-year strategic plan, the 2018 National Strategy for Preventing Veteran Suicide. This strategy adopts a comprehensive public health approach, integrating community-based and clinical interventions to effectively combat Veteran suicide[7].

Through an in-depth analysis of suicide ideation alerts via the Veterans Crisis Line's VSignals program, this paper aims to shed light on effective strategies for crisis outreach and provide valuable insights into utilizing

---

[1] Veteran's Crisis Line Public Fact Sheet (research.va.gov)
[1] VA Suicide Prevention Efforts, H.R. 4173 (cbo.gov)
[3] National Strategy for Preventing Veteran Suicide (va.gov),
[4] About Medallia (medallia.com)
[5] Policy_Brief_14_Caring_Letters.pdf (va.gov)
[6] CRS Veteran Suicide Prevention Report (congress.gov)
[7] National Strategy for Preventing Veteran Suicide (va.gov)

technological tools to understand the complete picture of the data from the trust surveys to leverage to make data-driven decisions for suicide prevention and crisis outreach.

Methods and Materials:

The analysis presented in this paper was conducted through a comprehensive process involving multiple stages to ensure accurate categorization and meaningful interpretation of respondent comments. Prior to analysis, a critical step involved safeguarding sensitive information. Personal Health Information (PHI) and Personally Identifiable Information (PII) were redacted from the dataset to uphold respondent confidentiality and adhere to data protection regulations. The methodology encompassed data collection, privacy protection measures, categorization utilizing OpenAI in Rows.com, and data analysis using Microsoft Excel.

The dataset included 3,607 respondent comments received during SERVE's involvement period from 8:00 am EST to January 13, 2023, to 10:00 pm EST on September 27, 2023. These comments were acquired through the VSignals program utilizing Medallia, a widely used platform for collecting and analyzing customer feedback. After all sensitive data was removed, the data set was standardized and formatted for data analysis using Microsoft Excel with the goal to explore and investigate gaps and trends in the data set.

Furthermore, analysts categorized each alert by utilizing pre-trained LLMs via Rows.com to classify relevant categories based on the content of the respondent comment. Categories included Care and Services, along with their respective subcategories, as well as Benefits and Other.

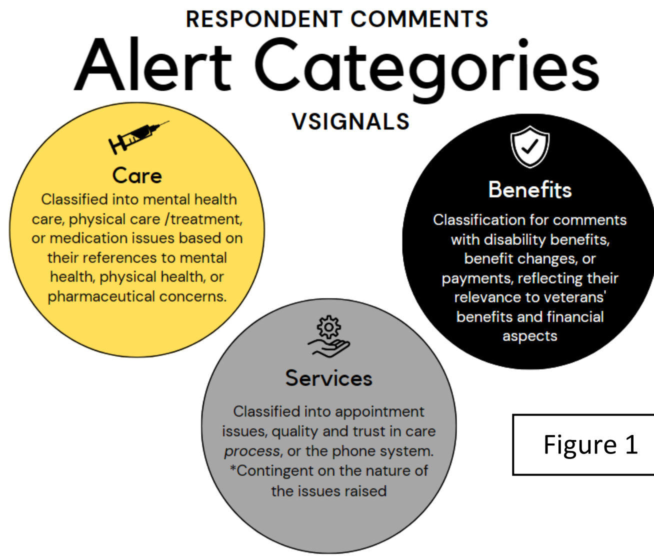


Figure 1

Our analysts calculated the total number of alerts falling within each category and subcategory to provide a quantitative overview of the dataset. Analysts assessed instances of overlapping categories to understand the potential complexity and nuances of the data and the potential for alerts to have relevance to multiple categories. The data analysis phase involved identifying patterns, trends, and recurring issues within each category and subcategory, shedding light on concerns raised by respondents. While this approach provided valuable insights, it is important to acknowledge that there are inherent limitations as the complexities of many comments may not have been fully captured by the assigned primary category. While LLM integrations were useful here for the purposes of demographic and trend analysis, we advise caution with using off-the-shelf, pre-trained LLMs to directly prioritize alert severity as VCL and VSignals address crisis situations for Veterans. Note that in this zero-shot classification setting, all LLM judgments are determined independently of one another using PHI and PIIscrubbed respondent comments, rather than the full supervision or support of the data, and without comprehensive knowledge of common overtures in patient outcomes.

In conclusion, this comprehensive methodology involved data collection, privacy protection, categorization with AI, and data analysis using Microsoft Excel. By categorizing and analyzing respondent comments and investigating trends, this inquiry aimed to gain valuable insights into veterans' experiences and concerns within the specified timeframe. The approach sought to strike a balance between thoroughness and the recognition of potential complexities inherent in qualitative data analysis.

Problem Statement:

In the context of SERVE's engagement from January 13, 2023, to September 27, 2023, analysis was completed, and it was revealed that many data values were missing, and the underutilization of Medallia is depicted in the visual below. The below table shows which survey category types (ie: Race/Ethnicity) have missing information and the prevalence (aggregate count, percentage) for the Alert Type: All Alerts (3607 alerts), True positive alerts (1560 alerts), and alerts which resulted in the acceptance of Suicide Prevention Coordinator (SPC) intervention (545 alerts). The prevalence of missing data is significant, limiting the conclusions that can be drawn from this data set.

Alerts with Missing Information

Survey Category	All Alerts	True Positive Alerts	SPC Intervention Accepted
Race/Ethnicity	3606, 100%	1560, 100%	545, 100%
Keywords Found	2954, 82%	1226, 79%	431, 79%
Race	1503, 42%	703, 45%	239, 44%
Ethnicity	1451, 40%	672, 43%	229, 42%
Population Density	934, 26%	512, 33%	186, 34%
Clinic Class	629, 17%	259, 17%	79, 14%
Clinic Name	594, 16%	259, 17%	79, 14%
Gender	524, 15%	210, 13%	72, 13%
Facility	416, 12%	161, 10%	50, 9%
Permission to Contact	257, 7%	144, 9%	47, 9%
VISN	214, 6%	119, 8%	36, 7%
Outcome	72, 2%	70, 4%	-
Comment Type	57, 2%	62, 4%	27, 5%
Respondent Comment*	55, 2%	25, 2%	10, 2%
Age	9, 0%	5, 0%	2, 0%

* These alerts also affect the ability to classify.

Table 1

Classification Insights: With the data provided, an examination has yielded the subsequent classifications for alerts: Care, Benefits, and Services. It is important to note that 55 Respondent Comments did not appear in the Medallia Report for unknown reasons and therefore, there are 55 alerts that were not classified.

For the **Care classification,** there exists a total of 2108 alerts from the following subcategories:

- Mental Health Care (797 alerts) *
    - Alerts that pertain to comments from respondents that specifically address mental health concerns, often of a critical nature, soliciting assistance, expressing thoughts of surrender, or referencing conditions such as PTSD, depression, or suicide.
- Physical Care and Treatment (987 alerts) *
    - Alerts that encompass remarks related to non-mental facets of care, such as physical discomfort, cardiovascular matters, and similar subjects.
- Medication Issues (324 alerts)
    - Alerts which relate to matters concerning pharmaceutical refills and other challenges related to medication.

*Mental and physical health care alerts simply describe the experience, whereas Quality and Trust in Care alerts, discussed later, provide opinions on those experiences.

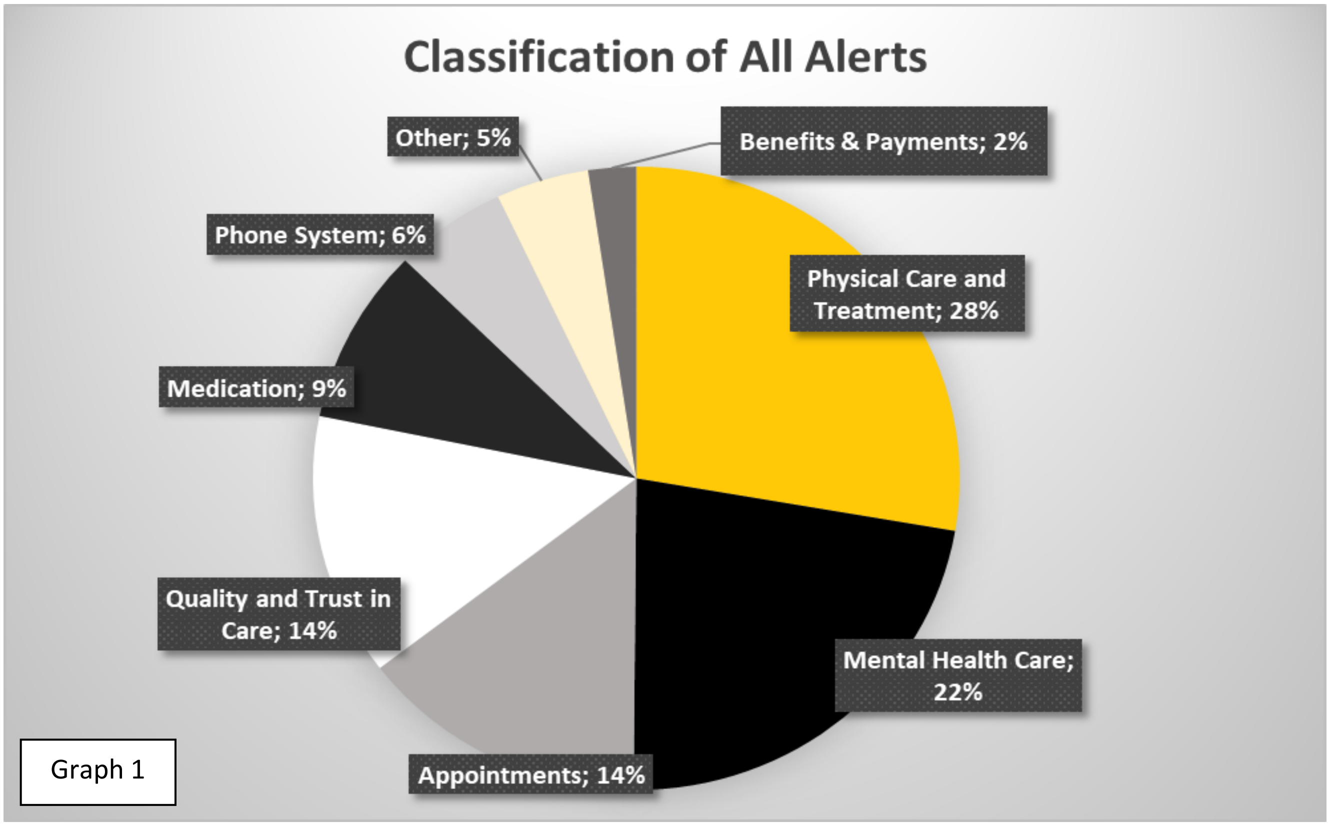
The **Services classification** comprises 1197 alerts from the following subcategories.

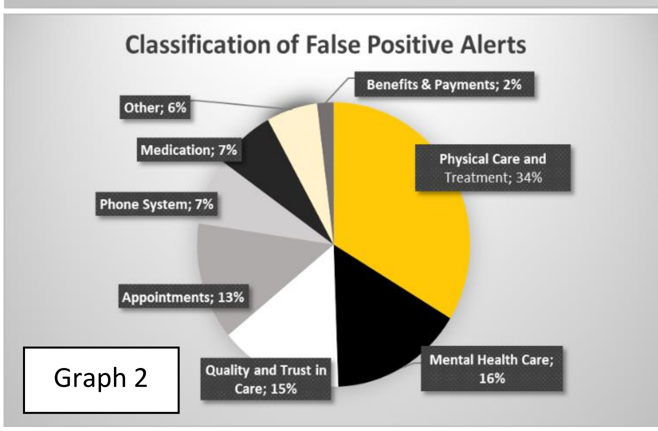
- Appointment Issues (513 alerts)
    - Alerts detailing challenges encountered while attempting to schedule appointments, concerns connected to telehealth, or delays in securing appointments with specialists, which could also include delays in referrals.
- Quality And Trust In Care (484 alerts)
    - Alerts that express a wide spectrum of viewpoints in both mental and physical health care, regarding opinions of care or provider experience. These comments may contain both positive and negative sentiments spanning from disagreement, dissatisfaction, or mishandling of care to appreciation, satisfaction, and trust in the care or provider.
- Phone System (200 alerts)
    - Alerts that address aspects such as extended wait times, the inclusion of suicide prevention messages during hold periods, and challenges faced while navigating the phone tree.

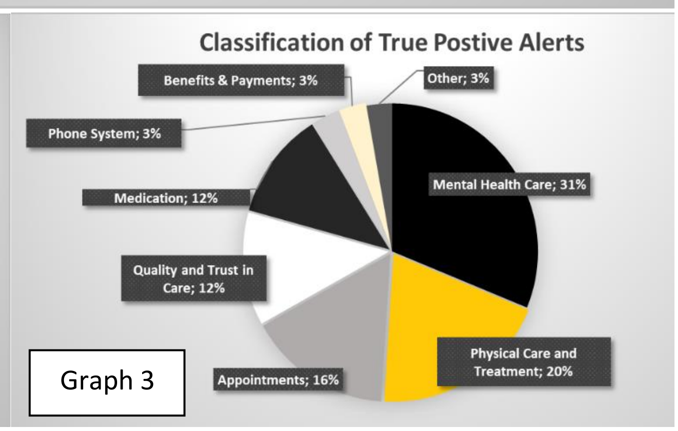
Within the **Benefits & Payments classification**, 85 alerts are documented. These alerts are related to eligibility, application, or other challenges related to disability benefits, changes in veteran benefits, or the financial aspects of care. Due to the small size of these alerts, subcategories were not created.

An **"Other" classification** category was created, and contains 165 alerts elucidating reasons for missed appointments, such as personal losses (e.g., "I missed my appointment due to a family bereavement"), complaints about COVID-19 mandates, and any other content that is not described above.

The classifications of all alerts, true positive alerts and false positive alerts are shown below.







Mental Health Care, Physical Health Care, and Appointments are in the top three classifications for all alerts and true positive alerts. False positive alerts also had Mental Health Care and Physical Health Care in the top two categories of alerts, however, Quality and Trust in Care was the third most classified. Additionally, Physical Care was more prevalent in false positive alerts compared to all alerts and true positive alerts.

Numerous alerts possess the potential to be classified under multiple categories. However, for this analysis, we report the primary zero-shot LLM classification for each alert. It is important to acknowledge that this approach does not comprehensively capture the complexity and intricacies present within veterans' comments. We leave assessments of comorbidity and orthogonal risk factors as future work, particularly for those authorized to access complete record-level and historical patient medical data. For our model prototyping and analysis, we analyze and transform the data as given to address potential areas of operational improvement. Note that while pragmatically useful, our mutually exclusive comment categorization remains incomplete: Respondent remarks frequently demonstrate a propensity for multiple categories of risk assessment, and the language used may convey a mixture of both affirmative and adverse sentiments. The following example illustrates this complexity:

"My problem is with the VA healthcare system at large. I feel frustrated that I am suffering everyday from some unknown condition without any medication to help. That I had referrals to see a specialist before but I had to start the process all over again. And in the meanwhile, I just have to figure it out by myself everyday. I can't work. I can barely function. I want to kill myself."

In this instance, a general sense of discontent with the healthcare system, challenges in accessing care, and concerns about physical and mental health are evident.

In the same process outlined above, respondent comments were classified into multiple categories, which produced the following results of the ten most occurring combinations of multiple classification. This is shown in the table below.

Multi-Classification	**Count**
Physical Care and Treatment, Quality and Trust in Care	423
Mental Health Care, Physical Care and Treatment	256
Physical Care and Treatment	181
Appointments, Physical Care and Treatment	174
Mental Health Care, Physical Care and Treatment, Quality and Trust in Care	139
Mental Health Care, Appointments	120
Quality and Trust in Care	98
Other	97
Medication, Quality and Trust in Care	85
Physical Care and Treatment, Medication, Quality and Trust in Care	76

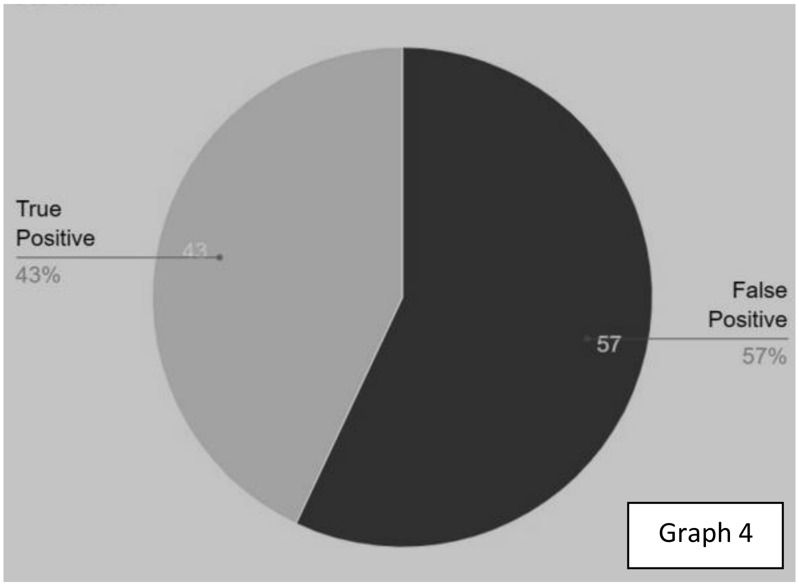
Table 2

This table shows Physical Care and Treatment and Quality of Trust in Care are most commo[illegible]as are Mental Health Care and Physical Care and Treatment, which suggests that these categories are closely related. Additional analysis using a comprehensive dataset holds the potential to pinpoint root causes and potential solutions relating to suicide prevention and outreach for Veterans.

Although we utilize proprietary LLMs via OpenAI integration with spreadsheet software to generate these classifications, we find it important to note here for process engineering purposes that pre-trained LLMs operate remotely. Remote parties may access sensitive information for quality control purposes and subject conversations to human review, which demanded that we exercise adequate cleaning of PII and PHI ahead of time to ensure compliance with the Health Insurance Portability and Accountability Act (HIPAA). This cleaning process may not be trivial to automate at scale, potentially necessitating the use of a locally operable alternative to cloud-based

LLMs-as-a-service. Additionally, LLMs' slow speed, their cost, their hallucinatory and sycophantic tendencies, and operators' inability to trivially explain their judgments in a black-box setting all pose grave operational risks in a mission-critical application such as directly predicting respondent proclivities for self-harm. Therefore, in the interest of compliance, expediency, and replicability, we limit ourselves to the use of open-access and open-source machine learning software for further modeling of patient outcomes conducted in this report, without LLM supervision or pseudo-annotation.

Alert Designation and Outcome Insights: For this line of inquiry, the Data Entry Operators marked alerts either as True or False Positives, based on their suicide-risk assessment of the respondent comments. Alerts that contained a risk of suicide or self-harm were sent to VCL supervisors, who conducted outreach and documented the outcomes. Alerts Marked as False Positives accounted for 2046 alerts (57%) and Alerts Marked as True Positives accounted for 1560 alerts (43%). One alert was not designated as a True Positive or False positive Alert, nor did it have an Outcome listed. Possibilities include a gap in data, or that this alert may not have been triaged or addressed appropriately.



Graph 4

Alerts Marked as False Positive typically have the outcome "Not a Valid Alert, No Outreach Attempted", which accounts for 2016 alerts (57% of total alerts, which we can define as the False Positive Rate, or FPR). However, there are thirty alerts in this data set where Alerts Marked as False Positive have an Outcome that differs from "Not a Valid Alert, No Outreach Attempted". This could be the result of human error in data entry, misinterpretation of risk of suicide ideation or self-harm, or other reasons. This highlights the need for standardization with automation to reduce discrepancies within the data set. The high False Positive rate also signifies a need for technological process improvement to reduce the number of false alerts coming in, which would in turn reduce time spent triaging alerts and free resources to focus on outreach efforts.

Alerts Marked as True Positive have one of five options based on outreach efforts and of these "Contact Made, Assessed - No Further Action" occurred 718 times (47%, which can be defined as the False Negative Rate, or FNR), "Contact Made, SPC Request Accepted" occurred 547 times (36%), "Outreach Attempted, No Contact Made – No Further Action" occurred 161 times (11%), "Contact made, Unable to assess/Veteran declined to engage" occurred 78 times (5%), and "Outreach Attempted, No Contact Made – Other Action" occurred 16 times (1%). There are 69 alerts, in addition to the aforementioned alert that do not have an outcome listed, which could be due to the open status of the alert as outreach is attempted or incomplete documentation of alerts.

The high FPR coupled with the high FNR signals a significant gap in both technological and human assessment of alerts that were flagged from Medallia as being related to suicidal ideation or self-harm. In total, there were 2734 alerts that did not require intervention; 2016 were alerts that that Medallia flagged in which human interpretation designated as False Positives and 718 Alerts Marked as True Positive by Medallia and human interpretation that did not require intervention, which suggest a human over-designation of True Positive alerts. This number could be higher if intervention was necessary, but not achievable as in the instances of the outcomes "Outreach Attempted, No Contact Made – No Further Action" and "Contact made, Unable to assess/Veteran declined to engage", which total 239 alerts, however obtaining that value is unachievable.

Therefore, based on the data set, only 15% of all alerts resulted in acceptance of a request for a Suicide Prevention Coordinator. This signifies that between and 84%-85% of all alerts do not require intervention, highlighting the need for refined technology and VCL processes, while bearing in mind the sensitive and intricate nature of the situation at hand as it relates to potential life and death situations for Veterans in crisis.

It may be necessary to consider potential algorithm adjustments within the Medallia platform to effectively address this divide.

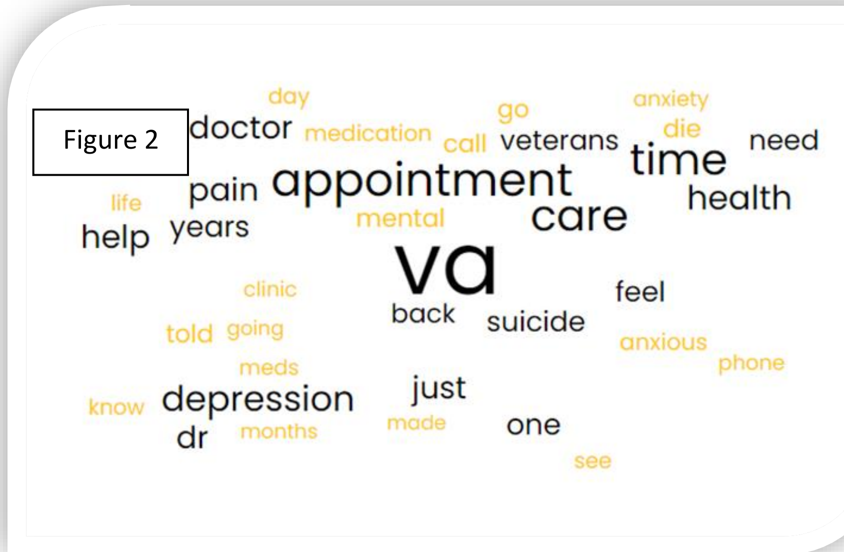
However, additional text analysis was conducted by analysts and revealed the 25 most common keywords in the entirety of the respondent comments as is listed in the table and graphic below. ("Doctor" and "Dr." were combined as were "medication" and "meds" as the latter are shortened word versions of the entirety of the preceding word). Looking at these frequently used words gives a glimpse into the common themes in the VSignals alert data, which could be used to enhance the accuracy of the alert system in flagging true positive alerts.

Word	Frequency in All Alerts
Doctor/Dr.	809
appointment	659
time	614
care	612
medication/meds	481
depression	470
health	439
help	417
pain	412
years	370
suicide	347
feel	346
need	342
veterans	336
mental	298

die	296
anxious	261
medication	250
life	240
anxiety	232
meds	231
clinic	226
phone	225

Table 3

Figure 2



Highlighting Key Concerns and Topics in Veterans' Healthcare Conversations

In summary, this paper has looked to categorize and analyze respondent comments within the specified period. It is imperative to acknowledge the intricate and complex nature of the respondent comments and their ability to intersect multiple categories, underscoring the need for a nuanced understanding when interpreting the data and the need for a comprehensive data set and additional utilization of technology to more accurately and efficiently categorize respondent comments to identify trends in veteran survey responses.

Solution:

The systematic categorization and analysis of respondent comments is important in unraveling valuable insights, trends, and concerns within the veterans' feedback, specifically for VSignals to identify suicide ideation trends and respond accordingly. However, we found that beyond analytical tool assistance, it is possible to use AI technologies to directly model the severity of patient responses. To enhance the accuracy of analysis, it is imperative that data from Medallia is comprehensive and includes often-missing demographic information. The lack thereof restricts the ability to appropriately make inferences based on the data trends, which could potentially inhibit the identification of populations that may require earlier intervention or a more targeted approach to suicidal and self-harm thoughts. Once there is more complete veteran survey response data, the integration of advanced technological solutions emerges as a solution to analyze this complex and nuanced data efficiently and accurately. Specifically, Medallia's Text Analytics platform has the capabilities to provide a redefined approach for alert categorization and analysis.

Medallia's Text Analytics platform, accessible at https://www.medallia.com/platform/text-analytics/, stands as a promising option in this endeavor, partially since the implementation of Medallia is already established in the

Department of Veteran Affairs and Veterans Crisis Line environments. This platform utilizes Natural Language Processing (NLP), sentiment analysis, and semantic understanding to accurately decode textual data. The platform contains the ability to automatically identify, classify, and categorize the nuanced themes, sentiments, and subcategories present within respondent comments[8]. This can be particularly useful for identifying gaps and trends amongst various demographics, which could provide insight into how to better provide targeted support for various subpopulations of Veterans.

By streamlining the categorization process and minimizing subjectivity, Medallia's Text Analytics empowers a more comprehensive approach to understanding Veterans' experiences. The platform's real-time monitoring capabilities provide a dynamic lens through which emergent trends can be promptly identified, enabling agile and informed decision-making to allocate resources and further reduce veteran suicide. Additionally, Medallia has the functionality to flag alerts and push notifications for alerts that have not been closed within 72 hours, which is the cut off time for attempting contact if contact has not already been made. By implementing this measure, and additional layer of protection would guarantee alerts are addressed in a timely manner, preventing potential oversight due to human error, which may have occurred with up to 69 alerts in this time period. Combining the strengths of human expertise and technological innovation would likely produce optimal results, which would enhance the alert categorization and analysis and lead to more informed and efficient decision-making.

If the culmination of the Medallia functionality cannot be enlisted, then the use of keyword flagging techniques can be further improved by stemming or lemmatizing to find each word's common stems or infinitive forms, which will increase recall by preventing false negatives for alternate conjugations of the same root word. This will allow for alert severity to be ranked by the size of the set intersections between flagged keywords and natural-language outpatient responses. During modeling experiments, SERVE was able to fit a classifier against outcome-derived, VA intervention labels using open-source gradient-boosted trees implementation *XGBoost,* on the task of predicting VA intervention based on tabular features using supervision provided by our analysts (per Alert Designation and Outcome Insights). Following the outcome binarization (false/true positive annotation), one-hot encoding of categorical features, cyclical encoding of temporal features and application of a 2x sample weight on true positives, it reached 0.97 AUC/ROC (Area Under the Curve of the Receiving Operator Characteristic), a 10% FPR, and a 7% FNR on a held-out, representative subsample of the data. The predictive power of machine learning models and their potential to aid in reprioritizing VA suicide interventions is not lost on us. Whether open-source tooling or Medallia is used, the potential of tool assistance to reduce burdens on personnel when identifying the severity of responses—and ultimately to impact patient outcomes and operational efficiency—are difficult to overstate.

Conclusion:

This paper employs methodology to categorize and analyze respondent comments from the Veterans Crisis Line's VSignals program. The findings underscore the complexities and nuances within veterans' feedback, with interwoven themes spanning care, services, and benefits. Through categorization, patterns, trends, and concerns revealed, shedding light on the challenges faced by veterans seeking assistance.

[8] https://www.medallia.com/wp-content/uploads/pdf/resources/Medallia-Text-Analytics-Brochure.pdf

The potential of advanced technological solutions, such as Medallia's Text Analytics platform, could revolutionize the process of alert categorization and analysis. was acknowledged and recommended. This tool offers the promise of efficient and accurate identification of themes, sentiments, and subcategories within respondent comments, complementing human expertise and enhancing the overall analysis process.

This paper highlights the importance of understanding veterans' experiences through in-depth analysis of their feedback. By integrating advanced technology, the Veterans Crisis Line can further its mission of providing timely and effective support to veterans in need.