**Supplementary Material**

**Clarification on Role of Algorithms and Methods**

In this section, we briefly summarize the role of certain algorithms and software used for this work:

* **SÉANCE:** A software package that transform presence and abundance of specific words within a corpus (i.e., bag of words) into more interpretable semantic variables representing specific psychosocial constructs. While future work is planned to interpret word frequency associations with suicide, dependent on specific patient characteristics, transformation was accomplished here to limit computational complexity by reducing the number of features and improve interpretation of the features. Fitting algorithms on all words present across the clinical notes would have made for more challenging interpretation and would have suffered from data sparsity.
* **SHAP Values:** SHAP values extend the shapley values concept from cooperative game theory, where, for model with features, the importance of feature is the weighted average of the marginal contribution of feature to the model’s prediction across all feature subsets:

Where is the subset of features excluding . The sum of the SHAP scores for a given datum from the expected value yields the model prediction– SHAP is an additive local model. SHAP can be calculated for individual datum, and the average of the absolute SHAP values yields the global feature importance. Tree SHAP can calculate SHAP values more efficient by leveraging the tree’s structure, weighting SHAP values by the number of training samples reaching specific leaves containing the feature, reducing the number of subsets that are considered to including those along the paths that the feature is used to split.

* **SHAP Interaction Scores:** Decomposes each SHAP value into its main effects and interactions ; . The calculation of via TreeSHAP is similar to that of the individual features, noting how the model prediction changes when considering tree paths where interacts with (i.e., decision split for , conditional on and vice versa). Predictions are compared with and versus without at least one of them across the relevant feature subsets by considering only relevant tree paths. Candidate interactions are identified by averaging across all datapoints, ranking global interaction contributions.
* **InteractionTransformer package:** Software developed to facilitate feature engineering of plausible interactions using the SHAP Interaction Scores. InteractionTransformer applies a CART model of choice to the data, calculates SHAP interaction scores, identifies and ranks their global feature importances, returning the top ranked terms and the new design matrix / engineered second order terms. This package returns candidate interactions () for further testing.
* **Calculation of p-value from candidate interactions to validate interactions:** Once the interactions have been identified, it is confirmed through statistical interaction testing, whereby the following model is fit on the note-level:   
  .   
  The p-value reflects the Wald statistic of the interaction term, :

Where .

* **Comparing Enrichment of Validated Interactions by Top SHAP Interactions:** We calculated the top 1000 ranked interactions by SHAP, assigned to set . Let represent the top 100 ranked interactions by SHAP, a subset of . Let represent the subset of interactions from that were validated through the method above. We hypothesize that truly validated interactions would be found predominantly in the set as opposed to the remaining interactions in . The Fisher’s exact calculates the odds ratio of the probability of finding in (with probability ) as opposed to set (with probability ). The odds ratio is defined as . The returned p-value represents enrichment of within .
* **Ridge Regression with validated interactions:** With the set of validated interactions, , we then fit the following model:

With the following penalty: . Here, and only refer to the select features/interactions themselves rather than observations.