

## Executive Summary-LARAN

Through an analysis of written transcriptions from U.S. presidential and vice-presidential debates, we make progress toward determining terminology and rhetoric that is more commonly used by the Democratic or Republican political parties. We used various NLP models such as Naive Bayes, random forests, deep-learning transformer models (BERT) for classification of text polarity over time.

Our project serves lobbyists with private interests, as well as politicians and speechwriters from both political parties. From preliminary polarization analysis, our clients can utilize several salient characteristics of the processed text.

For some issues, such as abortion, our model accurately reflects the severe polarization of our nation, along party lines. For some issues we identify a preference for specific terminology to label the same concept. For example, Democrats prefer “illegal immigrant” to “illegal alien” and further prefer “undocumented immigrant” to “illegal immigrant.” Finally, we find trends of certain terms occurring with increasing or decreasing probabilities over time, within the rhetoric of a single party. For example, our data analysis indicates that Democrats show an increasing preference for use of the term “white supremacy” in recent years. Our analysis assists politicians in constructing and communicating their campaign platforms, and it helps lobbyists to revise and polish their language to appeal to a target political party.

We have presented an effective model with potential to be refined and fortified by incorporating further analysis of structure and nuance in the English language. We would begin to accomplish this through more extensive data filtering and processing, to include feature reduction and word reduction. A next stage would incorporate intrinsic sentiment polarity into our analysis. We might then consider the yet unexplored problem of determining how polarization and sentiments are correlated with political approval ratings.