

## Meeting Notes 1/4/2025

Talking about Roberta and brainstorming methods to fine tune the code for our purposes.

Ryan: We might need to relabel our data to supervise it, since we'd have to make a classification algorithm.

Felix: we could give each word different scores ourselves and let the algorithm decide based on that, but it'd be biased.

How are we gonna make the machine values when it has nothing to work on in the first place? We need to give it default points to work with. Thus, we need to ask Voytek how to go about this.

Has knowledge of ML concepts but has no experience. Similarly with Minh.

Minh: How about we train it on datasets from previous papers to get their optimal parameters and then use our algorithm to work on our dataset. However, this poses the question of why we need to make our own model in the first place. → Do we collect more datasets from all sorts of papers and work with that? → The problem lies in how these papers collected the data and we'd need to tune the data or our algorithm to each of them, worst case. So it's not really practical.

Jonathan:

Homework: email professor Voytek about how to classify the words and quantify their political slants. Look more into how to modify Roberta.