Due to the large dataset imbalance, I employ techniques to reduce the impact of the class sizes on the model. Without this, the model would be heavily biased towards the Unrelated class in part 2.a and the Discuss class in part 2.b.

Care needs to be taken especially with respect to fake news because people may trust the machine learning models to correctly classify the relationship between the headlines and bodies without caring about the potential for false positives. As such, if some one with an ulterior motive, say political, publicised a model with knowingly poor classification (e.g. over classifying discuss instead of agree/disagree) people may take the information in the articles as fact without checking themselves because a model has classified it as such.