The SNLI corpus (version 1.0) is a collection of 570k human-written English sentence pairs manually labeled for balanced classification with the labels *entailment*, *contradiction*, and *neutral*, supporting the task of natural language inference (NLI), also known as recognizing textual entailment (RTE). We aim for it to serve both as a benchmark for evaluating representational systems for text, especially including those induced by representation learning methods, as well as a resource for developing NLP models of any kind.

### **Inspiration**

This dataset has been used to evaluate academic work on sentence encoding-based models for 3 way classification, with previous scores tabulated at<https://nlp.stanford.edu/projects/snli/>. Most of the entries use deep learning. How close to those scores (peak of 88.8% test accuracy) can you get with less computationally intensive methods?

### **Acknowledgements**

This dataset was kindly made available by the [Stanford Natural Language Processing Group](https://nlp.stanford.edu/). Please cite it as:

[Samuel R. Bowman](https://www.nyu.edu/projects/bowman/), [Gabor Angeli](http://cs.stanford.edu/~angeli/), [Christopher Potts](http://www.stanford.edu/~cgpotts/), and [Christopher D. Manning](http://nlp.stanford.edu/~manning/). 2015. *A large annotated corpus for learning natural language inference. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*