Datasets Sources

1. First dataset called 'Hello-SimpleAI/HC3' is obtained from https://huggingface.co/datasets/Hello-SimpleAI/HC3 using the python code that is used by Yannick Stephan.
2. Second dataset is obtained https://www.kaggle.com/competitions/human-or-machine-generated-text/data?select=train.tbz2
3. The third dataset is collected from Kaggle NaveenFream. (2023, September 22). Ai-and-human text. Kaggle. https://www.kaggle.com/datasets/naveenfream/ai-and-human-text