

## LIAR: A BENCHMARK DATASET FOR FAKE NEWS DETECTION

William Yang Wang, "Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection, to appear in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017), short paper, Vancouver, BC, Canada, July 30-August 4, ACL.

=====  
Description of the TSV format:

Column 1: the ID of the statement ([ID].json).  
Column 2: the label.  
Column 3: the statement.  
Column 4: the subject(s).  
Column 5: the speaker.  
Column 6: the speaker's job title.  
Column 7: the state info.  
Column 8: the party affiliation.  
Column 9-13: the total credit history count, including the current statement.  
9: barely true counts.  
10: false counts.  
11: half true counts.  
12: mostly true counts.  
13: pants on fire counts.  
Column 14: the context (venue / location of the speech or statement).

Note that we do not provide the full-text verdict report in this current version of the dataset,  
but you can use the following command to access the full verdict report and links to the source documents:  
wget [http://www.politifact.com//api/v/2/statement/\[ID\]/?format=json](http://www.politifact.com//api/v/2/statement/[ID]/?format=json)

=====  
The original sources retain the copyright of the data.

Note that there are absolutely no guarantees with this data,  
and we provide this dataset "as is",  
but you are welcome to report the issues of the preliminary version of this data.

You are allowed to use this dataset for research purposes only.

For more question about the dataset, please contact:  
William Wang, [william@cs.ucsb.edu](mailto:william@cs.ucsb.edu)

v1.0 04/23/2017