**Credibility Study 2019**

**Overview**

**DESCRIPTION**

This dataset of credibility annotations on over 2,000 environment/climate-related articles was created in 2019.

**SUMMARY**

This project contains dataset of over 2,000 articles annotated for credibility. These articles, which focus on the environment and science, are mostly annotated by 3 crowd raters each; a smaller subset of this data with more annotators per article and expert review can be found at <https://data.world/credibilitycoalition/credibility-factors2020/>.

Before participation, crowd raters filled out a demographic survey followed by committing to an Annotator Code of Conduct of performing their duties in as accurate and diligent a manner as possible provided in their informed consent. Once eligible, they received reading and rating tasks which included their credibility perception per article on a 5-point Likert scale, ranging from very low (1) to very high (5). Instructions to fill out the seven question survey across a 7-10 day period (estimated at 10 hours) were provided in a handbook with a recommended limit of 10-15 minutes per article. Articles chosen were written in English and represented a range of liberal to conservative positions or attitudes towards climate problems. To gather articles, the team used the Buzzsumo social media research tool in late 2018 to find the most popular articles over the previous year with the keywords of “climate change,” “global warming,” “environment,” and “pollution.” Among the top results, our team selected a set of articles with varying amounts of scientific reference.

As of yet, no academic publications have been written on this larger dataset. If you do any analysis or produce a publication, we would be excited to know.

* Questionnaire for crowd raters:
  + Credibility Coalition Study 2019 - Instructions - REDACTED.pdf
* Set of annotations
  + CredCo Study 2019 Crowd Annotators-FULL.csv

For 50 articles present in this dataset, demographic information related to raters was collected. Information about this can be found at <https://data.world/credibilitycoalition/credibility-factors2020/>.

**License**

This dataset is being released under CC-BY-NC. Requests for commercial use or feedback can be addressed to hello [at] credibilitycoalition [dot] org.

**Credibility-Factors2020**

**Overview**

**DESCRIPTION**

Analysis of this dataset referenced in C+J Conference 2020

**SUMMARY**

This dataset focuses on 50 articles about climate science, which were annotated completely by 49 students, 26 Upwork workers, 3 science and 3 journalism experts. (This data was gathered in 2019 as part of a larger dataset which can be found at <https://data.world/credibilitycoalition/2019-study>)

Before participation, crowd raters filled out a demographic survey followed by committing to an Annotator Code of Conduct of performing their duties in as accurate and diligent a manner as possible provided in their informed consent. Once eligible, they received reading and rating tasks that included their credibility perception per article on a 5-point Likert scale, ranging from very low (1) to very high (5). Instructions to fill out the seven question survey across a 7-10 day period (estimated at 10 hours) were provided in a handbook with a recommended limit of 10-15 minutes per article.

Articles chosen were written in English and represented a range of liberal to conservative positions or attitudes towards climate problems. To gather articles, the team used the Buzzsumo social media research tool in late 2018 to find the most popular articles over the previous year with the keywords of “climate change,” “global warming,” “environment,” and “pollution.” Among the top results, our team selected a set of articles with varying amounts of scientific reference.

More information about this dataset can be found within two papers, released or to be released in 2020:

1. Md Momen Bhuiyan, Amy X. Zhang, Connie Moon Sehat, and Tanushree Mitra. 2020. "Investigating 'Who' in the Crowdsourcing of News Credibility." In *Proceedings of Computation+Journalism Symposium (C+J’20)*. (pdf available at <https://credibilitycoalition.org/research/>)
2. Forthcoming: Md Momen Bhuiyan, Amy X. Zhang, Connie Moon Sehat, and Tanushree Mitra. 2020. "Investigating Differences in Crowdsourced News CredibilityAssessment: Raters, Tasks, and Expert Criteria." In *Proceedings of the ACM in Human-Computer Interaction.* (to be presented October 2020).

This project contains the following files:

* Questionnaire for crowd raters:
  + Credibility Coalition Study 2019 - Instructions - REDACTED.pdf
* Crowd annotator scoring
  + CredCo 2019 Study - CplusJ 2020 subset.csv
* List of Crowd Rater Annotator IDs, and crowd group [CredCo 'Student' or 'Upwork']
  + CredCo Study 2019 Crowd Annotators -simple.csv
* Expert scoring of the 50 articles
  + No Notes Copy of CredCo Study 2019 Expert Scores - Science #1.xlsx
  + No Notes Copy of CredCo Study 2019 Expert Scores - Science #2.xlsx
  + No Notes Copy of CredCo Study 2019 Expert Scores - Science #3.xlsx
  + No Notes Copy of CredCo Study 2019 Expert Scores - Journalism #A.xlsx
  + No Notes Copy of CredCo Study 2019 Expert Scores - Journalism #B.xlsx
  + No Notes Copy of CredCo Study 2019 Expert Scores - Journalism #C.xlsx

Not present in this dataset are two types of information:

* demographic related information of our anonymized crowd raters, including age, gender, education, political affiliation, economic/social issues attitudes, ethnicity, religion, English proficiency, and country of birth.
* qualitative comments by anonymized experts on the articles.

This information can be formally requested at [hello@credibilitycoalition.org.](mailto:hello@credibilitycoalition.org) Please identify yourself, the purposes of the request, the length of time you will plan on keeping the file before destroying it, and contact information of the data steward of the file (if yourself, please provide an email address and how long it will be valid for).

**License**

This dataset has been released under CC-BY-NC. Requests for commercial use or feedback can be addressed to hello [at] credibilitycoalition [dot] org.