NLP for Internet Freedom:
Censorship, Disinformation, and Propaganda

Proceedings of the Fourth Workshop

June 6, 2021
Copyright of each paper stays with the respective authors (or their employers).

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-26-8
Preface

Welcome to the fourth edition of the Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda. This is the second time we are running the workshop virtually, due to the COVID-19 pandemic. The pandemic has had a profound effect on the lives of people around the globe and, as much cliché as it may sound, we are all in this together. COVID-19 is the first pandemic in history in which technology and social media are being used on a massive scale to help people stay safe, informed, productive and connected. At the same time, the same technology that we rely on to keep us connected and informed is enabling and amplifying an infodemic that continues to undermine the global response and is detrimental to the efforts to control the pandemic. In the context of the pandemic, we define an infodemic as deliberate attempts to disseminate false information to undermine the public health response and to advance alternative agendas promoted by groups or individuals. We also realize that there is a vicious cycle: the more content is produced, the more misinformation and disinformation expand. The Internet is overloaded with false cures, such as drinking bleach or colloidal silver, conspiracy theories about the virus’ origin, false claims that the COVID-19 vaccine will change patient’s DNA or will serve to implant in us a tracking microchip, misinformation about methods such masks and social distancing, myths about the dangers and the real-world consequences of COVID-19. The goal of our workshop is to explore how we can address these issues using natural language processing.

We accepted 20 papers: 11 for the regular track and 9 for the shared tasks. The work presented at the workshop ranges from hate speech detection (Lemmens et al., Markov et al., Bose et al.) to approaches to identify false information and to verify facts (Maronikolakis et al., Bekoulis et al., Weld et al., Hämäläinen et al., Kazemi et al., Alhindi et al., Zuo et al.). The authors focus on different aspects of misinformation and disinformation: misleading headlines vs. rumor detection vs. stance detection for fact-checking. Some work concentrates on detecting biases in news media (e.g., Li & Goldwasser), which in turn contributes to research on propaganda. They all present different technical approaches.

Our workshop featured two shared tasks: Task 1 on Fighting the COVID-19 Infodemic, and Task 2 on Online Censorship Detection. Task 1 asked to predict several binary properties of a tweet about COVID-19, such as whether the tweet contains a verifiable factual claim or to what extent it is harmful to the society/community/individuals; the task was offered in multiple languages: English, Bulgarian, and Arabic. The second task was to predict whether a Sina Weibo tweet will be censored; it was offered in Chinese.

We are also thrilled to be able to bring two invited speakers: 1) Filippo Menczer, a distinguished professor of informatics and computer science at Indiana University, Bloomington, and Director of the Observatory on Social Media; and 2) Margaret Roberts, an associate professor of political science at University of California San Diego.

We thank the authors and the task participants for their interest in the workshop. We would also like to thank the program committee for their help with reviewing the papers and with advertising the workshop.

This material is partly based upon work supported by the US National Science Foundation under Grants No. 1704113 and No. 1828199.

It is also part of the Tanbih mega-project, which is developed at the Qatar Computing Research Institute, HBKU, and aims to limit the impact of “fake news,” propaganda, and media bias by making users aware of what they are reading, thus promoting media literacy and critical thinking.

The NLP4IF 2021 Organizers:

Anna Feldman, Giovanni Da San Martino, Chris Leberknight and Preslav Nakov
Organizing Committee

Anna Feldman, Montclair State University
Giovanni Da San Martino, University of Padova
Chris Leberknight, Montclair State University
Preslav Nakov, Qatar Computing Research Institute, HBKU

Program Committee

Firoj Alam, Qatar Computing Research Institute, HBKU
Tariq Alhindi, Columbia University
Jisun An, Singapore Management University
Chris Brew, LivePerson
Meghna Chaudhary, University of South Florida
Jedidah Crandall, Arizona State University
Anjalie Field, Carnegie-Mellon University
Parush Gera, University of South Florida
Yiqing Hua, Cornell University
Jeffrey Knockel, Citizen Lab
Haewoon Kwak, Singapore Management University
Verónica Pérez-Rosas, University of Michigan
Henrique Lopes Cardoso, University of Porto
Yelena Mejova, ISI Foundation
Jing Peng, Montclair State University
Marinella Petrocchi, IIT-CNR
Shaden Shaar, Qatar Computing Research Institute, HBKU
Matthew Sumpter, University of South Florida
Michele Tizzani, ISI Foundation
Brook Wu, New Jersey Institute of Technology
Table of Contents

Identifying Automatically Generated Headlines using Transformers
Antonis Maronikolakis, Hinrich Schütze and Mark Stevenson .................................................. 1

Improving Hate Speech Type and Target Detection with Hateful Metaphor Features
Jens Lemmens, Ilia Markov and Walter Daelemans ................................................................. 7

Improving Cross-Domain Hate Speech Detection by Reducing the False Positive Rate
Ilia Markov and Walter Daelemans .............................................................................................. 17

Understanding the Impact of Evidence-Aware Sentence Selection for Fact Checking
Giannis Bekoulis, Christina Papagiannopoulou and Nikos Deligiannis ................................. 23

Leveraging Community and Author Context to Explain the Performance and Bias of Text-Based Deception Detection Models
Galen Weld, Ellyn Ayton, Tim Althoff and Maria Glenski ......................................................... 29

Never guess what I heard... Rumor Detection in Finnish News: a Dataset and a Baseline
Mika Hämmäläinen, Khalid Alnajjar, Niko Partanen and Jack Rueter .................................. 39

Extractive and Abstractive Explanations for Fact-Checking and Evaluation of News
Ashkan Kazemi, Zehua Li, Verónica Pérez-Rosas and Rada Mihalcea .................................. 45

Generalisability of Topic Models in Cross-corpora Abusive Language Detection
Tulika Bose, Irina Illina and Dominique Fohr ........................................................................... 51

AraStance: A Multi-Country and Multi-Domain Dataset of Arabic Stance Detection for Fact Checking
Tariq Alhindi, Amal Alabdulkarim, Ali Alshehri, Muhammad Abdul-Mageed and Preslav Nakov 57

MEAN: Multi-head Entity Aware Attention Network for Political Perspective Detection in News Media
Chang Li and Dan Goldwasser ................................................................................................... 66

An Empirical Assessment of the Qualitative Aspects of Misinformation in Health News
Chaoyuan Zuo, Qi Zhang and Ritwik Banerjee ......................................................................... 76

Findings of the NLP4IF-2021 Shared Tasks on Fighting the COVID-19 Infodemic and Censorship Detection
Shaden Shaar, Firoj Alam, Giovanni Da San Martino, Alex Nikolov, Wajdi Zaghouani, Preslav Nakov and Anna Feldman ................................................................. 82

DamascusTeam at NLP4IF2021: Fighting the Arabic COVID-19 Infodemic on Twitter Using AraBERT
Ahmad Hussein, Nada Ghneim and Ammar Joukhadar .......................................................... 93

NARNIA at NLP4IF-2021: Identification of Misinformation in COVID-19 Tweets Using BERTweet
Ankit Kumar, Naman Jhunjhunwala, Raksha Agarwal and Niladri Chatterjee ......................... 99

R00 at NLP4IF-2021 Fighting COVID-19 Infodemic with Transformers and More Transformers
Ahmed Al-Qarqaz, Dia Abujaber and Malak Abdullah Abdullah ............................................. 104

Multi Output Learning using Task Wise Attention for Predicting Binary Properties of Tweets: Shared-Task-On-Fighting the COVID-19 Infodemic
Ayush Suhane and Shreyas Kowshik ....................................................................................... 110
iCompass at NLP4IF-2021–Fighting the COVID-19 Infodemic
Wassim Henia, Oumayma Rjab, Hatem Haddad and Chayma Fourati ................. 115

Fighting the COVID-19 Infodemic with a Holistic BERT Ensemble
Georgios Tziafas, Konstantinos Kogkalidis and Tommaso Caselli ......................... 119

Detecting Multilingual COVID-19 Misinformation on Social Media via Contextualized Embeddings
Subhadarshi Panda and Sarah Ita Levitan .................................................. 125

Transformers to Fight the COVID-19 Infodemic
Lasitha Uyangodage, Tharindu Ranasinghe and Hansi Hettiarachchi ...................... 130

Classification of Censored Tweets in Chinese Language using XLNet
Shaikh Sahil Ahmed and Anand Kumar M .................................................. 136
Workshop Program

*Identifying Automatically Generated Headlines using Transformers*
Antonis Maronikolakis, Hinrich Schütze and Mark Stevenson

*Improving Hate Speech Type and Target Detection with Hateful Metaphor Features*
Jens Lemmens, Ilia Markov and Walter Daelemans

*Improving Cross-Domain Hate Speech Detection by Reducing the False Positive Rate*
Ilia Markov and Walter Daelemans

*Understanding the Impact of Evidence-Aware Sentence Selection for Fact Checking*
Giannis Bekoulis, Christina Papagiannopoulou and Nikos Deligiannis

*Leveraging Community and Author Context to Explain the Performance and Bias of Text-Based Deception Detection Models*
Galen Weld, Ellyn Ayton, Tim Althoff and Maria Glenski

*Never guess what I heard... Rumor Detection in Finnish News: a Dataset and a Baseline*
Mika Hämäläinen, Khalid Alnajjar, Niko Partanen and Jack Rueter

*Extractive and Abstractive Explanations for Fact-Checking and Evaluation of News*
Ashkan Kazemi, Zehua Li, Verónica Pérez-Rosas and Rada Mihalcea

*Generalisability of Topic Models in Cross-corpora Abusive Language Detection*
Tulika Bose, Irina Illina and Dominique Fohr

*AraStance: A Multi-Country and Multi-Domain Dataset of Arabic Stance Detection for Fact Checking*
Tariq Alhindi, Amal Alabdulkarim, Ali Alshehri, Muhammad Abdul-Mageed and Preslav Nakov

*MEAN: Multi-head Entity Aware Attention Network for Political Perspective Detection in News Media*
Chang Li and Dan Goldwasser

*An Empirical Assessment of the Qualitative Aspects of Misinformation in Health News*
Chaoyuan Zuo, Qi Zhang and Ritwik Banerjee

*Findings of the NLP4IF-2021 Shared Tasks on Fighting the COVID-19 Infodemic and Censorship Detection*
Shaden Shaar, Firoj Alam, Giovanni Da San Martino, Alex Nikolov, Wajdi Zaghouani, Preslav Nakov and Anna Feldman
DamascusTeam at NLP4IF2021: Fighting the Arabic COVID-19 Infodemic on Twitter Using AraBERT
Ahmad Hussein, Nada Ghneim and Ammar Joukhadar

NARNIA at NLP4IF-2021: Identification of Misinformation in COVID-19 Tweets Using BERTweet
Ankit Kumar, Naman Jhunjhunwala, Raksha Agarwal and Niladri Chatterjee

R00 at NLP4IF-2021 Fighting COVID-19 Infodemic with Transformers and More Transformers
Ahmed Al-Qarqaz, Dia Abujaber and Malak Abdullah Abdullah

Multi Output Learning using Task Wise Attention for Predicting Binary Properties of Tweets: Shared-Task-On-Fighting the COVID-19 Infodemic
Ayush Suhane and Shreyas Kowshik

iCompass at NLP4IF-2021—Fighting the COVID-19 Infodemic
Wassim Henia, Oumayma Rjab, Hatem Haddad and Chayma Fourati

Fighting the COVID-19 Infodemic with a Holistic BERT Ensemble
Georgios Tziafas, Konstantinos Kogkalidis and Tommaso Caselli

Detecting Multilingual COVID-19 Misinformation on Social Media via Contextualized Embeddings
Subhadarshi Panda and Sarah Ita Levitan

Transformers to Fight the COVID-19 Infodemic
Lasitha Uyangodage, Tharindu Ranasinghe and Hansi Hettiarachchi

Classification of Censored Tweets in Chinese Language using XLNet
Shaikh Sahil Ahmed and Anand Kumar M.