

# Fact-Checking, Fake News, Propaganda, and Media Bias: Truth Seeking in the Post-Truth Era

**Preslav Nakov**

Qatar Computing Research Institute  
Hamad Bin Khalifa University  
pnakov@hbku.edu.qa

**Giovanni Da San Martino**

Department of Mathematics  
University of Padova  
dasan@math.unipd.it

## 1 Description

The rise of social media has democratized content creation and has made it easy for anybody to share and to spread information online. On the positive side, this has given rise to citizen journalism, thus enabling much faster dissemination of information compared to what was possible with newspapers, radio, and TV. On the negative side, stripping traditional media from their gate-keeping role has left the public unprotected against the spread of disinformation, which could now travel at breaking-news speed over the same democratic channel. This situation gave rise to the proliferation of false information specifically created to affect individual people's beliefs, and ultimately to influence major events such as political elections; it also set the dawn of the Post-Truth Era, where appeal to emotions has become more important than the truth. More recently, with the emergence of the COVID-19 pandemic, a new blending of medical and political misinformation and disinformation has given rise to the first global infodemic. Limiting the impact of these negative developments has become a major focus for journalists, social media companies, and regulatory authorities.

The tutorial offers an overview of the emerging and inter-connected research areas of fact-checking, misinformation, disinformation, “fake news”, propaganda, and media bias detection, with focus on text and on computational approaches. It further explores the general fact-checking pipeline and important elements thereof such as check-worthiness estimation, spotting previous fact-checked claims, stance detection, source reliability estimation, and detecting malicious users in social media. Finally, it covers some recent developments such as the emergence of large-scale pre-trained language models, and the challenges and opportunities they offer.

## 2 Outline of the Tutorial

Here is an outline of the tutorial. More information and materials are available online.<sup>1</sup>

### 2.1 Introduction

- (i) What is “fake news”?
  - (a) definitions
  - (b) properties
- (ii) “Fake news” as a weapon of mass deception
  - (a) impact on politics, finances, health
  - (b) can we win the war on “fake news”?

### 2.2 Check-Worthiness

- (i) Task definition
- (ii) Datasets
- (iii) Approaches
  - (a) ClaimBuster
  - (b) ClaimRank: modeling the context, multi-source learning, multi-linguality
  - (c) CLEF shared tasks

### 2.3 Detecting Previously Fact-Checked Claims

- (i) Task definition
- (ii) Datasets, e.g., CLEF
- (iii) Approaches

### 2.4 Fact-checking

- (i) Task definitions
- (ii) Datasets, e.g., Snopes, FEVER, ClamsKG, MultiFC
- (iii) Information sources: knowledge bases, Wikipedia, Web, social media, multimedia, tables

---

<sup>1</sup><http://propaganda.qcri.org/emnlp20-tutorial>

- (iv) Tasks and approaches
  - (a) fact-checking against knowledge bases
  - (b) fact-checking against Wikipedia
  - (c) fact-checking claims using the Web
  - (d) fact-checking rumors in social media
  - (e) fact-checking multi-modal claims, e.g., about images
  - (f) fact-checking the answers in community question answering forums
- (v) Shared tasks at SemEval and FEVER

## 2.5 Fake News Detection

- (i) Task definitions
- (ii) Datasets, e.g., FakeNewsNet, NELA-GT-2018
- (iii) The language of fake news
- (iv) Tasks and approaches

## 2.6 Stance Detection

- (i) Task definitions and examples
- (ii) Datasets
- (iii) Stance detection as a key element of fact-checking
- (iv) Information sources: text, social context, user profile
- (v) Tasks and approaches
  - (a) neural methods for stance detection
  - (b) cross-language stance detection
- (vi) Shared tasks at SemEval and the Fake News Challenge

## 2.7 Source Reliability and Media Bias Estimation

- (i) Task definitions and examples
- (ii) Datasets: Media Bias Fact/Check, AllSides, OpenSources, etc.
- (iii) Source reliability as a key element of fact-checking
- (iv) Special case: hyper-partisanship
- (v) Information sources: article text, Wikipedia, social media
- (vi) Tasks and approaches
  - (a) neural methods for source reliability estimation
  - (b) multi-modality
  - (c) multi-task learning

## 2.8 Propaganda Detection

- (i) Task definitions and examples
- (ii) Propaganda techniques and examples
- (iii) Datasets
- (iv) Tasks and approaches

## 2.9 Malicious User Detection

- (i) Typology of malicious users
- (ii) Datasets
- (iii) Tasks and approaches

## 2.10 Recent Developments and Future Challenges

- (i) Deep fakes: images, voice, video, text
- (ii) Text generation: GPT-2, GPT-3, GROVER
- (iii) Defending against neural fake news
- (iv) Fighting the COVID-19 Infodemic

## 3 Reading List

We recommend several surveys. [Shu et al. \(2017\)](#), which adopted a data mining perspective on “fake news” and focused on social media. Another survey ([Zubiaga et al., 2018a](#)) focused on rumor detection in social media. The survey by [Thorne and Vlachos \(2018\)](#) took a fact-checking perspective on “fake news” and related problems. The survey by [Li et al. \(2016\)](#) covering truth discovery in general. [Lazer et al. \(2018\)](#) offers a general overview and discussion on the science of “fake news”, while [Vosoughi et al. \(2018\)](#) focuses on the process of proliferation of true and false news online. Other recent surveys focus on stance detection ([Küçük and Can, 2020](#)), on propaganda ([Da San Martino et al., 2020b](#)), on social bots ([Ferrara et al., 2016](#)), on false information ([Zannettou et al., 2019b](#)) and on bias on the Web ([Baeza-Yates, 2018](#)).

See also the list of references at the end.

## 4 Type of Tutorial

The tutorial is both introductory, covering a number of topics related to fact-checking, propaganda and disinformation, but it is also cutting-edge, covering some latest developments in these areas.

## 5 Prerequisites

Prior knowledge of natural language processing, machine learning, and deep learning would be

needed in order to understand large parts of the contents of this tutorial.

## 6 Lecturers

### 6.1 Preslav Nakov

Dr. Preslav Nakov is a Principal Scientist at the Qatar Computing Research Institute (QCRI), HBKU. His research interests include computational linguistics, “fake news” detection, fact-checking, machine translation, question answering, sentiment analysis, lexical semantics, Web as a corpus, and biomedical text processing. He received his PhD degree from the University of California at Berkeley, and he was a Research Fellow in the National University of Singapore, a honorary lecturer in the Sofia University, and research staff at the Bulgarian Academy of Sciences.

At QCRI, he leads the Tanbih project,<sup>2</sup> developed in collaboration with MIT, which aims to limit the effect of “fake news”, propaganda and media bias by making users aware of what they are reading. The project was featured by over 100 news outlets, including Forbes, Boston Globe, Al-jazeera, MIT Technology Review, Science Daily, Popular Science, Fast Company, The Register, WIRED, and Engadget, among others.

As part of the project, he co-organized several shared tasks on fact-checking and propaganda detection at SemEval and CLEF, as well as a related NLP4IF workshop.

He is President of ACL SIGLEX, a Secretary of ACL SIGSLAV, and a member of the EACL advisory board. He is also member of the editorial board of TACL, CS&L, NLE, AI Communications, and Frontiers in AI, as well as of the Language Science Press Book Series on Phraseology and Multiword Expressions. He co-authored a Morgan & Claypool book on Semantic Relations between Nominals, two books on computer algorithms, and many research papers in top-tier conferences and journals. He received the Young Researcher Award at RANLP’2011, and he was the first to receive the Bulgarian President’s John Atanasoff award, named after the inventor of the first automatic electronic digital computer.

### 6.2 Giovanni Da San Martino

Giovanni Da San Martino is a Senior Assistant Professor at the University of Padova, Italy. His research interests are at the intersection of machine learning and natural language processing. He has been researching for 10+ years on these topics, publishing more than 60 publications in top-tier conferences and journals. He received his PhD from the University of Bologna, he was a Research Fellow at the University of Padova and a Scientist at Qatar Computing Research Institute. He has worked on several NLP tasks including paraphrase recognition, stance detection and community question answering. Currently, he is actively involved in researching on disinformation and propaganda detection. As part of this research he has been co-organiser of the Checkthat! labs at CLEF 2018-2020, the NLP4IF 2019-2020 workshops on “censorship, disinformation, and propaganda”, the 2019 Hack the News Datathon and the SemEval-2020 task 11 on “Detection of propaganda techniques in news articles.”

---

<sup>2</sup>Tanbih project: <http://tanbih.qcri.org>

## References

- Aseel Addawood, Adam Badawy, Kristina Lerman, and Emilio Ferrara. 2019. Linguistic cues to deception: Identifying political trolls on social media. In *Proceedings of the International AAAI Conference on Web and Social Media, ICWSM '19*, pages 15–25, Munich, Germany.
- Firoj Alam, Fahim Dalvi, Shaden Shaar, Nadir Durrani, Hamdy Mubarak, Alex Nikolov, Giovanni Da San Martino, Ahmed Abdelali, Hassan Sajjad, Kareem Darwish, and Preslav Nakov. 2020a. Fighting the COVID-19 infodemic in social media: A holistic perspective and a call to arms. *ArXiv preprint 2007.07996*.
- Firoj Alam, Shaden Shaar, Fahim Dalvi, Hassan Sajjad, Alex Nikolov, Hamdy Mubarak, Giovanni Da San Martino, Ahmed Abdelali, Nadir Durrani, Kareem Darwish, and Preslav Nakov. 2020b. Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, policy makers, and the society. *ArXiv preprint 2005.00033*.
- Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2):211–36.
- Jisun An, Meeyoung Cha, Krishna Gummadi, Jon Crowcroft, and Daniele Quercia. 2012. Visualizing media bias through Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media, ICWSM '12*, pages 2–5, Dublin, Ireland.
- Atanas Atanasov, Gianmarco De Francisci Morales, and Preslav Nakov. 2019. Predicting the role of political trolls in social media. In *Proceedings of the 2019 SIGNLL Conference on Computational Natural Language Learning, CoNLL '19*, pages 1023–1034, Hong Kong, China.
- Pepa Atanasova, Lluís Màrquez, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Wajdi Zaghouani, Spas Kyuchukov, Giovanni Da San Martino, and Preslav Nakov. 2018. Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims, Task 1: Check-worthiness. In *CLEF 2018 Working Notes. Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France. CEUR-WS.org.
- Pepa Atanasova, Preslav Nakov, Georgi Karadzhov, Mitra Mohtarami, and Giovanni Da San Martino. 2019a. Overview of the CLEF-2019 CheckThat! Lab on Automatic Identification and Verification of Claims. Task 1: Check-Worthiness. In *CLEF 2019 Working Notes. Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Lugano, Switzerland. CEUR-WS.org.
- Pepa Atanasova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Georgi Karadzhov, Tsvetomila Mihaylova, Mitra Mohtarami, and James Glass. 2019b. Automatic fact-checking using context and discourse information. *J. Data and Information Quality*, 11(3):12:1–12:27.
- Pepa Atanasova, Jakob Grue Simonsen, Christina Lioma, and Isabelle Augenstein. 2020. Generating fact checking explanations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL '20*, pages 7352–7364.
- Isabelle Augenstein, Christina Lioma, Dongsheng Wang, Lucas Chaves Lima, Casper Hansen, Christian Hansen, and Jakob Grue Simonsen. 2019. MultiFC: A real-world multi-domain dataset for evidence-based fact checking of claims. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP '19*, pages 4685–4697, Hong Kong, China.
- Mouhamadou Lamine Ba, Laure Berti-Equille, Kushal Shah, and Hossam M. Hammady. 2016. VERA: A platform for veracity estimation over web data. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16*, pages 159–162, Montréal, Québec, Canada.
- Adam Badawy, Kristina Lerman, and Emilio Ferrara. 2019. Who falls for online political manipulation? In *Companion Proceedings of The 2019 World Wide Web Conference, WWW '19*, page 162–168, San Francisco, California, USA.
- Ricardo Baeza-Yates. 2018. Bias on the web. *Commun. ACM*, 61(6):54–61.
- Ramy Baly, Georgi Karadzhov, Dimitar Alexandrov, James Glass, and Preslav Nakov. 2018a. Predicting factuality of reporting and bias of news media sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP '18*, pages 3528–3539, Brussels, Belgium.
- Ramy Baly, Georgi Karadzhov, Jisun An, Haewoon Kwak, Yoan Dinkov, Ahmed Ali, James Glass, and Preslav Nakov. 2020. What was written vs. who read it: News media profiling using text analysis and social media context. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL '20*, pages 3364–3374.
- Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James Glass, and Preslav Nakov. 2019. Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT '19*, pages 2109–2116, Minneapolis, Minnesota, USA.

- Ramy Baly, Mitra Mohtarami, James Glass, Lluís Màrquez, Alessandro Moschitti, and Preslav Nakov. 2018b. Integrating stance detection and fact checking in a unified corpus. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT '18*, pages 21–27, New Orleans, Louisiana, USA.
- Alberto Barrón-Cedeño, Giovanni Da San Martino, Israa Jaradat, and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5):1849 – 1864.
- Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Lluís Màrquez, Pepa Atanasova, Wajdi Zaghouni, Spas Kyuchukov, Giovanni Da San Martino, and Preslav Nakov. 2018. Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims, Task 2: Factuality. In *CLEF 2018 Working Notes. Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, CEUR Workshop Proceedings, Avignon, France. CEUR-WS.org.
- Alberto Barrón-Cedeño, Giovanni Da San Martino, Israa Jaradat, and Preslav Nakov. 2019. Proppy: A system to unmask propaganda in online news. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, AAAI'19*, pages 9847–9848, Honolulu, HI, USA.
- Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, and Fatima Haouari. 2020. Check-That! at CLEF 2020: Enabling the automatic identification and verification of claims in social media. In *Advances in Information Retrieval, ECIR '20*, pages 499–507, Lisbon, Portugal.
- Alberto Barrón-Cedeño, Tamer Elsayed, Preslav Nakov, Giovanni Da San Martino, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Nikolay Babulkov, Bayan Hamdan, Alex Nikolov, Shaden Shaar, and Zien Sheikh Ali. 2020. Overview of CheckThat! 2020 — automatic identification and verification of claims in social media. In *Proceedings of the 11th International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction, CLEF '2020*, Thessaloniki, Greece.
- Anya Belz. 2019. Fully automatic journalism: We need to talk about nonfake news generation. In *Proceedings of the First Conference on Truth and Trust Online, TTO '19*, London, United Kingdom.
- Gillian Bolsover and Philip Howard. 2017. Computational propaganda and political big data: Moving toward a more critical research agenda. *Big Data*, 5(4):273–276.
- Zhan Bu, Zhengyou Xia, and Jiandong Wang. 2013. A sock puppet detection algorithm on virtual spaces. *Know.-Based Syst.*, 37:366–377.
- Ceren Budak, Sharad Goel, and Justin M Rao. 2016. Fair and balanced? Quantifying media bias through crowdsourced content analysis. *Public Opinion Quarterly*, 80(S1):250–271.
- Kevin R. Canini, Bongwon Suh, and Peter L. Pirolli. 2011. Finding credible information sources in social networks based on content and social structure. In *Proceedings of the IEEE International Conference on Privacy, Security, Risk, and Trust, and the IEEE International Conference on Social Computing, SocialCom/PASSAT '11*, pages 1–8, Boston, Massachusetts, USA.
- Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on Twitter. In *Proceedings of the 20th International Conference on World Wide Web, WWW '11*, page 675–684, Hyderabad, India.
- Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2013. Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5):560–588.
- Abhijnan Chakraborty, Bhargavi Paranjape, Kakarla Kakarla, and Niloy Ganguly. 2016. Stop clickbait: Detecting and preventing clickbaits in online news media. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '16*, pages 9–16, San Francisco, California, USA.
- Cheng Chen, Kui Wu, Venkatesh Srinivasan, and Xudong Zhang. 2013. Battling the Internet Water Army: detection of hidden paid posters. In *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '13*, pages 116–120, Niagara, Canada.
- Sihao Chen, Daniel Khashabi, Wenpeng Yin, Chris Callison-Burch, and Dan Roth. 2019a. Seeing things from a different angle: discovering diverse perspectives about claims. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT '19*, pages 542–557, Minneapolis, Minnesota, USA.
- Wenhu Chen, Hongmin Wang, Jianshu Chen, Yunkai Zhang, Hong Wang, Shiyang Li, Xiyong Zhou, and William Yang Wang. 2019b. TabFact: A large-scale dataset for table-based fact verification. *arXiv preprint 1909.02164*.
- Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoni, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. The COVID-19 social media infodemic. *arXiv preprint 2003.05004*.
- Michael Conover, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization

- on Twitter. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, ICWSM '11, pages 89–96, Barcelona, Spain.
- Giovanni Da San Martino, Alberto Barrón-Cedeño, Henning Wachsmuth, Rostislav Petrov, and Preslav Nakov. 2020a. SemEval-2020 task 11: Detection of propaganda techniques in news articles. In *Proceedings of the International Workshop on Semantic Evaluation*, SemEval '20, Barcelona, Spain.
- Giovanni Da San Martino, Alberto Barron-Cedeno, and Preslav Nakov. 2019a. Findings of the NLP4IF-2019 shared task on fine-grained propaganda detection. In *Proceedings of the 2nd Workshop on NLP for Internet Freedom (NLP4IF): Censorship, Disinformation, and Propaganda*, NLP4IFEMNLP '19, pages 162–170, Hong Kong, China.
- Giovanni Da San Martino, Stefano Cresci, Alberto Barrón-Cedeño, Seunghak Yu, Roberto Di Pietro, and Preslav Nakov. 2020b. A survey on computational propaganda detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI-PRICAI '20, pages 4826–4832.
- Giovanni Da San Martino, Shaden Shaar, Yifan Zhang, Seunghak Yu, Alberto Barrón-Cedeño, and Preslav Nakov. 2020. Prta: A system to support the analysis of propaganda techniques in the news. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, ACL '20, pages 287–293.
- Giovanni Da San Martino, Seunghak Yu, Alberto Barron-Cedeno, Rostislav Petrov, and Preslav Nakov. 2019b. Fine-grained analysis of propaganda in news articles. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, EMNLP '19, pages 5636–5646, Hong Kong, China.
- Kareem Darwish, Dimitar Alexandrov, Preslav Nakov, and Yelena Mejova. 2017. Seminar users in the Arabic Twitter sphere. In *Proceedings of the 9th International Conference on Social Informatics*, SocInfo '17, pages 91–108, Oxford, UK.
- Kareem Darwish, Michael Aupetit, Peter Stefanov, and Preslav Nakov. 2020. Unsupervised user stance detection on Twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM '20, pages 141–152, Atlanta, Georgia, USA.
- Sohan De Sarkar, Fan Yang, and Arjun Mukherjee. 2018. Attending sentences to detect satirical fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, COLING '18, pages 3371–3380, Santa Fe, New Mexico, USA.
- Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559.
- Stefano DellaVigna and Ethan Kaplan. 2007. The Fox News effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.
- Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga. 2017. SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours. In *Proceedings of the 11th International Workshop on Semantic Evaluation*, SemEval '17, pages 60–67, Vancouver, Canada.
- Yoan Dinkov, Ahmed Ali, Ivan Koychev, and Preslav Nakov. 2019a. Predicting the leading political ideology of YouTube channels using acoustic, textual, and metadata information. In *Proceedings of the 20th Annual Conference of the International Speech Communication Association*, INTERSPEECH '19, pages 501–505, Graz, Austria.
- Yoan Dinkov, Ivan Koychev, and Preslav Nakov. 2019b. Detecting toxicity in news articles: Application to Bulgarian. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, RANLP '19, pages 247–258.
- Ming Dong, Bolong Zheng, Nguyen Quoc Viet Hung, Han Su, and Guohui Li. 2019. Multiple rumor source detection with graph convolutional networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM '19, page 569–578, Beijing, China.
- Xin Luna Dong, Evgeniy Gabrilovich, Kevin Murphy, Van Dang, Wilko Horn, Camillo Lugaresi, Shao-hua Sun, and Wei Zhang. 2015. Knowledge-based trust: Estimating the trustworthiness of web sources. *Proc. VLDB Endow.*, 8(9):938–949.
- Sebastian Dungs, Ahmet Aker, Norbert Fuhr, and Kalina Bontcheva. 2018. Can rumour stance alone predict veracity? In *Proceedings of the 27th International Conference on Computational Linguistics*, COLING '18, pages 3360–3370, Santa Fe, New Mexico, USA.
- Erick Elejalde, Leo Ferres, and Eelco Herder. 2018. On the nature of real and perceived bias in the mainstream media. *PloS one*, 13(3):e0193765.
- Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Pepa Atanasova, and Giovanni Da San Martino. 2019a. CheckThat! at CLEF 2019: Automatic identification and verification of claims. In *Proceedings of the 41st European Conference on Information Retrieval*, ECIR '19, pages 309–315, Cologne, Germany.
- Tamer Elsayed, Preslav Nakov, Alberto Barrón-Cedeño, Maram Hasanain, Reem Suwaileh, Giovanni Da San Martino, and Pepa Atanasova. 2019b.

- Overview of the CLEF-2019 CheckThat!: Automatic identification and verification of claims. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, LNCS, Lugano, Switzerland. Springer.
- Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Commun. ACM*, 59(7):96–104.
- William Ferreira and Andreas Vlachos. 2016. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT '16*, pages 1163–1168, San Diego, California, USA.
- Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public opinion quarterly*, 80(S1):298–320.
- Mohamed H. Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. 2019a. ExFaKT: A framework for explaining facts over knowledge graphs and text. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, WSDM '19*, pages 87–95, Melbourne, Australia.
- Mohamed H. Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. 2019b. Tracy: Tracing facts over knowledge graphs and text. In *The World Wide Web Conference, WWW '19*, pages 3516–3520, San Francisco, California, USA.
- Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the International World Wide Web Conference, WWW '18*, pages 913–922, Lyon, France.
- Sebastian Gehrmann, Hendrik Strobelt, and Alexander Rush. 2019. GLTR: Statistical detection and visualization of generated text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, ACL '19*, pages 111–116, Florence, Italy.
- Pepa Gencheva, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. 2017. A context-aware approach for detecting worth-checking claims in political debates. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP '17*, pages 267–276, Varna, Bulgaria.
- Matthew Gentzkow and Jesse M Shapiro. 2006. Media bias and reputation. *Journal of political Economy*, 114(2):280–316.
- Genevieve Gorrell, Ahmet Aker, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2019. SemEval-2019 task 7: RumourEval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 845–854, Minneapolis, Minnesota, USA.
- Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The Quarterly Journal of Economics*, 120(4):1191–1237.
- Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. Before name-calling: Dynamics and triggers of ad hominem fallacies in web argumentation. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT '18*, pages 386–396, New Orleans, Louisiana, USA.
- Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. 2018. A retrospective analysis of the fake news challenge stance-detection task. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING '18*, pages 1859–1874, Santa Fe, New Mexico, USA.
- Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In search of credible news. In *Proceedings of the 17th International Conference on Artificial Intelligence: Methodology, Systems, and Applications, AIMSA '16*, pages 172–180, Varna, Bulgaria.
- Maram Hasanain, Fatima Haouari, Reem Suwaileh, Zien Sheikh Ali, Bayan Hamdan, Tamer Elsayed, Alberto Barrón-Cedeño, Giovanni Da San Martino, and Preslav Nakov. 2020. Overview of the CLEF-2020 CheckThat! lab on automatic identification and verification of claims in social media: Arabic tasks. In *Working Notes of CLEF 2020—Conference and Labs of the Evaluation Forum, CLEF '2020*, Thessaloniki, Greece.
- Maram Hasanain, Reem Suwaileh, Tamer Elsayed, Alberto Barrón-Cedeño, and Preslav Nakov. 2019. Overview of the CLEF-2019 CheckThat! Lab on Automatic Identification and Verification of Claims. Task 2: Evidence and Factuality. In *CLEF 2019 Working Notes. Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Lugano, Switzerland. CEUR-WS.org*.
- Naeemul Hassan, Chengkai Li, and Mark Tremayne. 2015. Detecting check-worthy factual claims in presidential debates. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management, CIKM '15*, page 1835–1838, Melbourne, Australia.
- Naeemul Hassan, Afroza Sultana, You Wu, Gensheng Zhang, Chengkai Li, Jun Yang, and Cong Yu. 2014. Data in, fact out: Automated monitoring of facts by FactWatcher. *PVLDB*, 7:1557–1560.



- Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarini, Anil Kumar Nayak, Vikas Sable, Chengkai Li, and Mark Tremayne. 2017. ClaimBuster: The first-ever end-to-end fact-checking system. *Proc. VLDB Endow.*, 10(12):1945–1948.
- Christopher Hidey, Tuhin Chakrabarty, Tariq Alhindi, Siddharth Varia, Kriste Krstovski, Mona Diab, and Smaranda Muresan. 2020. DeSePtion: Dual sequence prediction and adversarial examples for improved fact-checking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL ’20, pages 8593–8606.
- Renee Hobbs and Sandra Mcgee. 2008. Teaching about propaganda: An examination of the historical roots of media literacy. *Journal of Media Literacy Education*, 6(62):56–67.
- Benjamin D. Horne, William Dron, Sara Khedr, and Sibel Adali. 2018a. Assessing the news landscape: A multi-module toolkit for evaluating the credibility of news. In *Proceedings of the The Web Conference*, WWW ’18, pages 235–238, Lyon, France.
- Benjamin D. Horne, Sara Khedr, and Sibel Adali. 2018b. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Proceedings of the Twelfth International Conference on Web and Social Media*, ICWSM ’18, pages 518–527, Stanford, California, USA.
- Benjamin D. Horne, Dorit Nevo, John O’Donovan, Jin-Hee Cho, and Sibel Adali. 2019a. Rating reliability and bias in news articles: Does AI assistance help everyone? In *Proceedings of the Thirteenth International Conference on Web and Social Media*, ICWSM ’19, pages 247–256, Munich, Germany.
- Benjamin D Horne, Jeppe Nørregaard, and Sibel Adali. 2019b. Different spirals of sameness: A study of content sharing in mainstream and alternative media. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM ’19, pages 257–266, Munich, Germany.
- Viet-Phi Huynh and Paolo Papotti. 2019a. A benchmark for fact checking algorithms built on knowledge bases. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM ’19, pages 689–698, Beijing, China.
- Viet-Phi Huynh and Paolo Papotti. 2019b. A benchmark for fact checking algorithms built on knowledge bases. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, CIKM ’19, page 689–698, Beijing, China.
- Viet-Phi Huynh and Paolo Papotti. 2019c. Buckle: Evaluating fact checking algorithms built on knowledge bases. *Proc. VLDB Endow.*, 12(12):1798–1801.
- Shanto Iyengar and Kyu S Hahn. 2009. Red media, blue media: Evidence of ideological selectivity in media use. *Journal of communication*, 59(1):19–39.
- Israa Jaradat, Pepa Gencheva, Alberto Barrón-Cedeño, Lluís Màrquez, and Preslav Nakov. 2018. Claim-Rank: Detecting check-worthy claims in Arabic and English. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT ’18, pages 26–30, New Orleans, Louisiana, USA.
- Shan Jiang, Miriam Metzger, Andrew Flanagin, and Christo Wilson. 2020. Modeling and measuring expressed (dis)belief in (mis)information. In *Proceedings of the International AAAI Conference on Web and Social Media*, ICWSM ’20, pages 315–326, Atlanta, Georgia, USA.
- Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. 2014. News credibility evaluation on microblog with a hierarchical propagation model. In *Proceedings of the 2014 IEEE International Conference on Data Mining*, ICDM 2014, ICDM ’14, pages 230–239, Shenzhen, China.
- Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI ’16, pages 2972–2978, Phoenix, AZ, USA.
- Georgi Karadzhov, Pepa Gencheva, Preslav Nakov, and Ivan Koychev. 2017a. We built a fake news & clickbait filter: What happened next will blow your mind! In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, RANLP ’17, pages 334–343, Varna, Bulgaria.
- Georgi Karadzhov, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, and Ivan Koychev. 2017b. Fully automated fact checking using external sources. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, RANLP ’17, pages 344–353, Varna, Bulgaria.
- Georgios Karagiannis, Immanuel Trummer, Saehan Jo, Shubham Khandelwal, Xuezhi Wang, and Cong Yu. 2019. Mining an “anti-knowledge base” from Wikipedia updates with applications to fact checking and beyond. *Proc. VLDB Endow.*, 13(4):561–573.
- Twin Karmakharm, Nikolaos Aletras, and Kalina Bontcheva. 2019. Journalist-in-the-loop: Continuous learning as a service for rumour analysis. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing: System Demonstrations*,



- EMNLP-IJCNLP '19, pages 115–120, Hong Kong, China.
- Dhruv Khattar, Jaipal Singh Goud, Manish Gupta, and Vasudeva Varma. 2019. MVAE: Multimodal variational autoencoder for fake news detection. In *Proceedings of the World Wide Web Conference, WWW '19*, page 2915–2921, San Francisco, California, USA.
- Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. Semeval-2019 task 4: Hyperpartisan news detection. In *Proceedings of the 13th International Workshop on Semantic Evaluation, SemEval '19*, pages 829–839, Minneapolis, Minnesota, USA.
- Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING '18*, pages 3402–3413, Santa Fe, New Mexico, USA.
- Lev Konstantinovskiy, Oliver Price, Mevan Babakar, and Arkaitz Zubiaga. 2018. Towards automated factchecking: Developing an annotation schema and benchmark for consistent automated claim detection. *ArXiv preprint 1809.08193*.
- Daniel Kopev, Ahmed Ali, Ivan Koychev, and Preslav Nakov. 2019. Detecting deception in political debates using acoustic and textual features. In *Proceedings of the IEEE Automatic Speech Recognition and Understanding Workshop, ASRU '19*, pages 652–659, Singapore.
- Dilek Küçük and Fazli Can. 2020. Stance detection: A survey. *ACM Comput. Surv.*, 53(1).
- Vivek Kulkarni, Junting Ye, Steve Skiena, and William Yang Wang. 2018. Multi-view models for political ideology detection of news articles. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP '18*, pages 3518–3527, Brussels, Belgium.
- Srijan Kumar, Justin Cheng, Jure Leskovec, and V.S. Subrahmanian. 2017. An army of me: sockpuppets in online discussion communities. In *Proceedings of the 26th International Conference on World Wide Web, WWW '17*, pages 857–866, Perth, Australia.
- Sejeong Kwon, Meeyoung Cha, and Kyomin Jung. 2017. Rumor detection over varying time windows. *PLoS ONE*, 12(1):e0168344.
- Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In *Proceedings of the 13th IEEE International Conference on Data Mining, ICDM '13*, pages 1103–1108, Dallas, Texas, USA.
- David M.J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. 2018. The science of fake news. *Science*, 359(6380):1094–1096.
- Yaliang Li, Jing Gao, Chuishi Meng, Qi Li, Lu Su, Bo Zhao, Wei Fan, and Jiawei Han. 2016. A survey on truth discovery. *SIGKDD Explor. Newsl.*, 17(2):1–16.
- Yang Liu and Yi-Fang Brook Wu. 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, AAAI '18*, pages 354–361, New Orleans, Louisiana, USA.
- Yi-Ju Lu and Cheng-Te Li. 2020. GCAN: Graph-aware co-attention networks for explainable fake news detection on social media. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL '20*, pages 505–514.
- Luca Luceri, Silvia Giordano, and Emilio Ferrara. 2020. Detecting troll behavior via inverse reinforcement learning: A case study of Russian trolls in the 2016 US election. In *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media, ICWSM '20*, pages 417–427, Atlanta, Georgia, USA.
- Michal Lukasik, Trevor Cohn, and Kalina Bontcheva. 2015. Point process modelling of rumour dynamics in social media. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, ACL-IJCNLP '15*, pages 518–523, Beijing, China.
- Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI '16*, pages 3818–3824, New York, NY, USA.
- Jing Ma, Wei Gao, and Kam-Fai Wong. 2017. Detect rumors in microblog posts using propagation structure via kernel learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL '17*, pages 708–717, Vancouver, Canada.
- Todor Mihaylov, Georgi Georgiev, and Preslav Nakov. 2015a. Finding opinion manipulation trolls in news community forums. In *Proceedings of the Nineteenth Conference on Computational Natural Language Learning, CoNLL '15*, pages 310–314, Beijing, China.

- Todor Mihaylov, Ivan Koychev, Georgi Georgiev, and Preslav Nakov. 2015b. Exposing paid opinion manipulation trolls. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, RANLP '15, pages 443–450, Hissar, Bulgaria.
- Todor Mihaylov, Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Georgi Georgiev, and Ivan Koychev. 2018. The dark side of news community forums: Opinion manipulation trolls. *Internet Research*, 28(5):1292–1312.
- Todor Mihaylov and Preslav Nakov. 2016. Hunting for troll comments in news community forums. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, ACL '16, pages 399–405, Berlin, Germany.
- Tsvetomila Mihaylova, Georgi Karadzhov, Pepa Atanasova, Ramy Baly, Mitra Mohtarami, and Preslav Nakov. 2019. SemEval-2019 task 8: Fact checking in community question answering forums. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, SemEval '19, pages 860–869, Minneapolis, Minnesota, USA.
- Tsvetomila Mihaylova, Preslav Nakov, Lluís Màrquez, Alberto Barrón-Cedeño, Mitra Mohtarami, Georgi Karadjov, and James Glass. 2018. Fact checking in community forums. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, AAAI '18, pages 5309–5316, New Orleans, Louisiana, USA.
- Sebastião Miranda, David Nogueira, Afonso Mendes, Andreas Vlachos, Andrew Secker, Rebecca Garrett, Jeff Mitchel, and Zita Marinho. 2019. Automated fact checking in the news room. In *Proceedings of the World Wide Web Conference*, WWW '19, pages 3579–3583, San Francisco, California, USA.
- Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry. 2016. SemEval-2016 task 6: Detecting stance in tweets. In *Proceedings of the 10th International Workshop on Semantic Evaluation*, SemEval '16, pages 31–41, San Diego, California, USA.
- Mitra Mohtarami, Ramy Baly, James Glass, Preslav Nakov, Lluís Màrquez, and Alessandro Moschitti. 2018. Automatic stance detection using end-to-end memory networks. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT '18, pages 767–776, New Orleans, Louisiana, USA.
- Mitra Mohtarami, James Glass, and Preslav Nakov. 2019. Contrastive language adaptation for cross-lingual stance detection. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, EMNLP '19, pages 4442–4452, Hong Kong, China.
- Meredith Ringel Morris, Scott Counts, Asta Roseway, Aaron Hoff, and Julia Schwarz. 2012. Tweeting is believing?: Understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW '12, pages 441–450, Seattle, Washington, USA.
- Subhabrata Mukherjee and Gerhard Weikum. 2015. Leveraging joint interactions for credibility analysis in news communities. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 353–362, Melbourne, Australia.
- Moin Nadeem, Wei Fang, Brian Xu, Mitra Mohtarami, and James Glass. 2019. FAKTA: An automatic end-to-end fact checking system. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, NAACL-HLT '19, pages 78–83, Minneapolis, Minnesota, USA.
- Preslav Nakov. 2020. Can we spot the “fake news” before it was even written? *arXiv preprint 2008.04374*.
- Preslav Nakov, Alberto Barrón-Cedeño, Tamer Elsayed, Reem Suwaileh, Lluís Màrquez, Wajdi Zaghouani, Pepa Atanasova, Spas Kyuchukov, and Giovanni Da San Martino. 2018. Overview of the CLEF-2018 CheckThat! Lab on automatic identification and verification of political claims. In *Proceedings of the Ninth International Conference of the CLEF Association: Experimental IR Meets Multilinguality, Multimodality, and Interaction*, Lecture Notes in Computer Science, pages 372–387, Avignon, France. Springer.
- Preslav Nakov, Tsvetomila Mihaylova, Lluís Màrquez, Yashkumar Shiroya, and Ivan Koychev. 2017. Do not trust the trolls: Predicting credibility in community question answering forums. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing*, RANLP '17, pages 551–560, Varna, Bulgaria.
- An T. Nguyen, Aditya Kharosekar, Matthew Lease, and Byron C. Wallace. 2018. An interpretable joint graphical model for fact-checking from crowds. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, AAAI '18, New Orleans, Louisiana, USA.
- Van-Hoang Nguyen, Kazunari Sugiyama, Preslav Nakov, and Min-Yen Kan. 2020. FANG: Leveraging social context for fake news detection using graph representation. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management*, CIKM '20.
- Yixin Nie, Haonan Chen, and Mohit Bansal. 2019. Combining fact extraction and verification with neural semantic matching networks. In *Proceedings of*

- the Thirty-Third AAAI Conference on Artificial Intelligence*, AAAI '19, pages 6859–6866, Honolulu, Hawaii, USA.
- Jeppe Nørregaard, Benjamin D. Horne, and Sibel Adali. 2019. NELA-GT-2018: A large multi-labelled news dataset for the study of misinformation in news articles. In *Proceedings of the Thirteenth International Conference on Web and Social Media*, ICWSM '19, pages 630–638, Munich, Germany.
- Symeon Papadopoulos, Kalina Bontcheva, Eva Jaho, Mihai Lupu, and Carlos Castillo. 2016. Overview of the special issue on trust and veracity of information in social media. *ACM Trans. Inf. Syst.*, 34(3):14:1–14:5.
- Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics*, COLING '18, pages 3391–3401, Santa Fe, New Mexico, USA.
- Evaggelia Pitoura, Panayiotis Tsaparas, Giorgos Flouris, Iirini Fundulaki, Panagiotis Papadakis, Serge Abiteboul, and Gerhard Weikum. 2018. On measuring bias in online information. *SIGMOD Rec.*, 46(4):16–21.
- Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2016. Credibility assessment of textual claims on the web. In *Proceedings of the 25th ACM International Conference on Information and Knowledge Management*, CIKM '16, pages 2173–2178, Indianapolis, Indiana, USA.
- Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the truth lies: Explaining the credibility of emerging claims on the Web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*, WWW '17, pages 1003–1012, Perth, Australia.
- Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2018. CredEye: A credibility lens for analyzing and explaining misinformation. In *Proceedings of The Web Conference 2018*, WWW '18, pages 155–158, Lyon, France.
- Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2018. A stylistic inquiry into hyperpartisan and fake news. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, ACL '18, pages 231–240, Melbourne, Australia.
- Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi. 2017. Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, EMNLP '17, pages 2931–2937, Copenhagen, Denmark.
- Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. 2011. Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference Companion on World Wide Web*, WWW '11, pages 249–252, Hyderabad, India.
- Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*, ACL '13, pages 1650–1659, Sofia, Bulgaria.
- Filipe N Ribeiro, Lucas Henrique, Fabricio Benvenuto, Abhijnan Chakraborty, Juhi Kulshrestha, Mahmoudreza Babaei, and Krishna P Gummadi. 2018. Media bias monitor: Quantifying biases of social media news outlets at large-scale. In *Proceedings of the Twelfth International AAAI Conference on Web and Social Media*, ICWSM '18, pages 290–299, Stanford, California, USA.
- Benjamin Riedel, Isabelle Augenstein, Georgios P Spithourakis, and Sebastian Riedel. 2017. A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. *ArXiv preprint 1707.03264*.
- Natali Ruchansky, Sungyong Seo, and Yan Liu. 2017. CSI: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM Conference on Information and Knowledge Management*, CIKM '17, pages 797–806, Singapore.
- Diego Saez-Trumper, Carlos Castillo, and Mounia Lalmas. 2013. Social media news communities: Gatekeeping, coverage, and statement bias. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*, CIKM '13, pages 1679–1684, San Francisco, California, USA.
- Shaden Shaar, Nikolay Babulkov, Giovanni Da San Martino, and Preslav Nakov. 2020a. That is a known lie: Detecting previously fact-checked claims. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL '20, pages 3607–3618.
- Shaden Shaar, Alex Nikolov, Nikolay Babulkov, Firoj Alam, Alberto Barrón-Cedeño, Tamer Elsayed, Maram Hasanain, Reem Suwaileh, Fatima Haouari, Giovanni Da San Martino, and Preslav Nakov. 2020b. Overview of the CLEF-2020 CheckThat! lab on automatic identification and verification of claims in social media: English tasks. In *Working Notes of CLEF 2020—Conference and Labs of the Evaluation Forum*, CLEF '2020, Thessaloniki, Greece.
- Gautam Kishore Shahi, Anne Dirksen, and Tim A. Maehrzak. 2020. An exploratory study of COVID-19 misinformation on Twitter. *ArXiv preprint 2005.05710*.

- Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion*, pages 745–750, Geneva, Switzerland.
- Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. 2017. Finding streams in knowledge graphs to support fact checking. In *Proceedings of the IEEE International Conference on Data Mining, ICDM '17*, New Orleans, Louisiana, USA.
- Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A data repository with news content, social context and dynamic information for studying fake news on social media. *arXiv preprint arXiv:1809.01286*.
- Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *SIGKDD Explor. Newsl.*, 19(1):22–36.
- Kai Shu, Suhang Wang, and Huan Liu. 2019. Beyond news contents: The role of social context for fake news detection. In *Proceedings of the 12th ACM International Conference on Web Search and Data Mining, WSDM '19*, pages 312–320, Melbourne, Australia.
- Peter Stefanov, Kareem Darwish, Atanas Atanasov, and Preslav Nakov. 2020. Predicting the topical stance and political leaning of media using tweets. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL '20*, pages 527–537.
- Edson C. Tandoc Jr., Zheng Wei Lim, and Richard Ling. 2018. Defining “fake news”. *Digital Journalism*, 6(2):137–153.
- Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. 2019. ClaimsKG: A knowledge graph of fact-checked claims. In *Proceedings of the 18th International Semantic Web Conference, ISWC '19*, pages 309–324, Auckland, New Zealand.
- Denis Teyssou, Jean-Michel Leung, Evlampios Apostolidis, Konstantinos Apostolidis, Symeon Papadopoulos, Markos Zampoglou, Olga Papadopoulou, and Vasileios Mezaris. 2017. The InVID plug-in: Web video verification on the browser. In *Proceedings of the First International Workshop on Multimedia Verification, MuVer '17*, pages 23–30, Mountain View, California, USA.
- James Thorne, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang, and Andreas Vlachos. 2017. Fake news stance detection using stacked ensemble of classifiers. In *Proceedings of the EMNLP Workshop on Natural Language Processing meets Journalism*, pages 80–83, Copenhagen, Denmark.
- James Thorne and Andreas Vlachos. 2018. Automated fact checking: Task formulations, methods and future directions. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING '18*, pages 3346–3359, Santa Fe, New Mexico, USA.
- James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018a. FEVER: a large-scale dataset for fact extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT '18*, pages 809–819, New Orleans, Louisiana, USA.
- James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2019a. Evaluating adversarial attacks against multiple fact verification systems. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP '19*, pages 2944–2953, Hong Kong, China.
- James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. 2018b. The fact extraction and VERification (FEVER) shared task. In *Proceedings of the First Workshop on Fact Extraction and VERification, FEVER '18*, pages 1–9, Brussels, Belgium.
- James Thorne, Andreas Vlachos, Oana Cocarascu, Christos Christodoulopoulos, and Arpit Mittal. 2019b. The FEVER2.0 shared task. In *Proceedings of the Second Workshop on Fact Extraction and VERification, FEVER '19*, pages 1–6, Hong Kong, China.
- Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. In *Proceedings of the Eleventh International Conference on Web and Social Media, ICWSM '17*, pages 280–289, Montréal, Québec, Canada.
- Slavena Vasileva, Pepa Atanasova, Lluís Màrquez, Alberto Barrón-Cedeño, and Preslav Nakov. 2019. It takes nine to smell a rat: Neural multi-task learning for check-worthiness prediction. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP '19*, pages 1229–1239, Varna, Bulgaria.
- Andreas Vlachos and Sebastian Riedel. 2014. Fact checking: Task definition and dataset construction. In *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, Maryland, USA.

- Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science*, 359(6380):1146–1151.
- William Yang Wang. 2017. “Liar, Liar Pants on Fire”: A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, ACL ’17, pages 422–426, Vancouver, Canada.
- Yaqing Wang, Fenglong Ma, Zhiwei Jin, Ye Yuan, Guangxu Xun, Kishlay Jha, Lu Su, and Jing Gao. 2018. EANN: Event adversarial neural networks for multi-modal fake news detection. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’18, page 849–857, London, United Kingdom.
- Lianwei Wu, Yuan Rao, Yongqiang Zhao, Hao Liang, and Ambreen Nazir. 2020. DTCA: Decision tree-based co-attention networks for explainable claim verification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL ’20, pages 1024–1035.
- You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, and Cong Yu. 2017. Computational fact checking through query perturbations. *ACM Trans. Database Syst.*, 42(1).
- Kai-Cheng Yang, Onur Varol, Clayton A. Davis, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2019. Arming the public with artificial intelligence to counter social bots. *Human Behavior and Emerging Technologies*, 1(1):48–61.
- Kai-Cheng Yang, Onur Varol, Pik-Mai Hui, and Filippo Menczer. 2020. Scalable and generalizable social bot detection through data selection. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence*, AAAI ’20, pages 1096–1103, New York, New York, USA.
- Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. 2019a. Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *Companion Proceedings of The 2019 World Wide Web Conference*, WWW ’19, page 218–226, San Francisco, California, USA.
- Savvas Zannettou, Michael Sirivianos, Jeremy Blackburn, and Nicolas Kourtellis. 2019b. The web of false information: Rumors, fake news, hoaxes, clickbait, and various other shenanigans. *J. Data and Information Quality*, 11(3):10:1–10:37.
- Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. In *Advances in Neural Information Processing Systems 32*, NeurIPS ’19, pages 9054–9065, Vancouver, Canada.
- Amy X. Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B. Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, Ed Bice, Sandro Hawke, David R. Karger, and An Xiao Mina. 2018. A structured response to misinformation: Defining and annotating credibility indicators in news articles. In *Proceedings of the The Web Conference*, WWW ’18 companion, pages 603–612, Lyon, France.
- Yifan Zhang, Giovanni Da San Martino, Alberto Barrón-Cedeño, Salvatore Romeo, Jisun An, Hae-woon Kwak, Todor Staykovski, Israa Jaradat, Georgi Karadzhov, Ramy Baly, Kareem Darwish, and Preslav Nakov James Glass. 2019. Tanbih: Get to know what you are reading. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, EMNLP ’19, pages 223–228, Hong Kong, China.
- Wanjuan Zhong, Jingjing Xu, Duyu Tang, Zenan Xu, Nan Duan, Ming Zhou, Jiahai Wang, and Jian Yin. 2020. Reasoning over semantic-level graph for fact checking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, ACL ’20, pages 6170–6180, Online.
- Xinyi Zhou, Apurva Mulay, Emilio Ferrara, and Reza Zafarani. 2020. ReCOVeRY: A multimodal repository for COVID-19 news credibility research. *arXiv preprint 2006.05557*.
- Dimitrina Zlatkova, Preslav Nakov, and Ivan Koychev. 2019. Fact-checking meets fauxtography: Verifying claims about images. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP ’19, pages 2099–2108, Hong Kong, China.
- Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018a. Detection and resolution of rumours in social media: A survey. *ACM Comput. Surv.*, 51(2).
- Arkaitz Zubiaga and Heng Ji. 2014. Tweet, but verify: epistemic study of information verification on Twitter. *Social Network Analysis and Mining*, 4(1):1–12.
- Arkaitz Zubiaga, Elena Kochkina, Maria Liakata, Rob Procter, Michal Lukasik, Kalina Bontcheva, Trevor Cohn, and Isabelle Augenstein. 2018b. Discourse-aware rumour stance classification in social media using sequential classifiers. *Inf. Process. Manage.*, 54(2):273–290.
- Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PLoS ONE*, 11(3):1–29.