MAD '22 Workshop: Multimedia AI against Disinformation

Conference Paper · June 2022

DOI: 10.1145/3512527.3531440

5 authors, including:

Bogdan Ionescu
Polytechnic University of Bucharest
220 PUBLICATIONS **2,439** CITATIONS
SEE PROFILE

Giorgos Kordopatis-Zilos
Czech Technical University in Prague
40 PUBLICATIONS **432** CITATIONS
SEE PROFILE

Adrian Popescu
Atomic Energy and Alternative Energies Commission
129 PUBLICATIONS **2,069** CITATIONS
SEE PROFILE

Luca Cuccovillo
Fraunhofer Institute for Digital Media Technology IDMT
25 PUBLICATIONS **135** CITATIONS
SEE PROFILE

MAD '22 Workshop: Multimedia AI against Disinformation

Bogdan Ionescu
Politehnica University of Bucharest
Bucharest, Romania
bogdan.ionescu@upb.ro

Giorgos Kordopatis-Zilos
ITI-CERTH
Thessaloniki, Greece
georgekordopatis@iti.gr

Adrian Popescu
Université Paris-Saclay, CEA, List
F-91120, Palaiseau, France
adrian.popescu@cea.fr

Luca Cuccovillo
Fraunhofer IDMT
Ilmenau, Germany
luca.cuccovillo@idmt.fraunhofer.de

Symeon Papadopoulos
ITI-CERTH
Thessaloniki, Greece
papadop@iti.gr

ABSTRACT

The verification of multimedia content posted online becomes increasingly challenging due to recent advancements in synthetic media manipulation and generation. Moreover, malicious actors can easily exploit AI technologies to spread disinformation across social media at a rapid pace, which poses very high risks for society and democracy. There is, therefore, an urgent need for AI-powered tools that facilitate the media verification process. The objective of the MAD '22 workshop is to bring together those who work on the broader topic of disinformation detection in multimedia in order to share their experiences and discuss their novel ideas, reaching out to people with different backgrounds and expertise. The research domains of interest vary from the detection of manipulated and synthetic content in multimedia to the analysis of the spread of disinformation and its impact on society. The MAD '22 workshop proceedings are available at: https://dl.acm.org/citation.cfm?id=3512732.

CCS CONCEPTS

• **Computing methodologies** → **Artificial intelligence**; • **Information systems** → **Multimedia information systems**.

KEYWORDS

disinformation detection, fake news, deepfakes, multimedia forensics, bot detection, audio forensics, fact-checking

ACM Reference Format:
Bogdan Ionescu, Giorgos Kordopatis-Zilos, Adrian Popescu, Luca Cuccovillo, and Symeon Papadopoulos. 2022. MAD '22 Workshop: Multimedia AI against Disinformation. In *Proceedings of the 2022 International Conference on Multimedia Retrieval (ICMR '22), June 27–30, 2022, Newark, NJ, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3512527.3531440

1 INTRODUCTION

Disinformation spreads rapidly in online social networks and is propagated by social media actors and network communities to achieve specific (mostly malevolent) objectives. Disinformation has deleterious effects on users' real lives since it distorts their points of view regarding societally-sensitive topics, such as politics, health or religion. Ultimately, it has a negative effect on the very fabric of democratic societies and should be fought against via an effective combination of human and technical means.

Disinformation campaigns are increasingly powered by advanced AI tools, and a lot of effort has been put into the detection of misleading and fake content. Yet, whether a piece of information is considered misleading or true often depends on the temporal and cultural contexts in which it is interpreted. This is, for instance, the case for scientific knowledge, which evolves at a fast pace, and whose usage in mainstream content should be updated accordingly.

Multimedia content is often at the core of AI-assisted disinformation campaigns. Their impact is directly related to the perceived credibility of their content. Due to the recent advances in deep learning, state-of-the-art algorithms can generate photo-realistic content, synthesize realistic speech, and compose human-like text that is very difficult to distinguish from authentic content. Also, combining these advances has the potential to boost the impact of disinformation. This combination is an open research topic that needs to be addressed in order to reduce its effects. The MAD '22 workshop aspires to partly address this open research area of increasing societal importance and, ultimately, contribute to the mitigation of disinformation.

2 SCOPE

Topics of interest for MAD '22 include but are not limited to:

- Disinformation detection in multimedia content
- Multimodal verification methods
- Synthetic, deepfake and manipulated media detection
- Multimedia forensics
- Disinformation spread and impact in social media
- Analysis of disinformation campaigns in societally-sensitive domains (e.g., politics, health)
- Explaining disinformation to non-experts
- Disinformation detection technologies for non-expert users
- Dataset sharing and governance in AI for disinformation
- Temporal and cultural aspects of disinformation
- Datasets for disinformation detection and media verification
- Multimedia verification systems and applications

3 OBJECTIVE

The aim of the first workshop on Multimedia AI against Disinformation (MAD '22) is to bring together people working on the topic

ICMR '22, June 27–30, 2022, Newark, NJ, USA

ACM ISBN 978-1-4503-9238-9/22/06.
https://doi.org/10.1145/3512527.3531440

of disinformation in relation to multimedia content. The workshop acts as a bridge to share their experiences and work in the field and discuss their brave new ideas. This will accelerate the advancements in the field of multimodal disinformation detection by providing the means for the communication of people working on different aspects of the problem, e.g. working with different modalities for the detection of fake content, and with different expertise, e.g. AI researchers, social scientists and fact-checkers.

4 INVITED AND KEYNOTE SPEAKERS

Prof. Katja de Vries is an Assistant Professor in public law at Uppsala University funded by the Ragnar Söderberg Foundation. Also, she is affiliated to the Swedish Law and Informatics Research Institute (Stockholm), the Center for Law, Science, Technology and Society (Brussels), and the Department of Sociology of Law (Lund). Her current research focuses on the challenges that AI-generated content ('deepfakes' or 'synthetic data') poses to data protection, intellectual property and other fields of law.

Prof. Luisa Verdoliva is an Associate Professor at University Federico II of Naples, Italy, where she leads the Multimedia Forensics Lab. In 2018, she has been visiting professor at Friedrich-Alexander-University (FAU) and in 2019-2020 she has been visiting scientist at Google AI in San Francisco. Her scientific interests are in the field of image and video processing, with main contributions in the area of multimedia forensics.

5 ACCEPTED PAPERS

The workshop accepted 10 submissions tackling various facets of multimedia AI techniques, especially for analyzing news, social media information, and deepfakes.

Bogaert et al. [2] present a study of the exploitation of language generation models for disinformation purposes, addressing two perspectives: (i) quantitatively, where language models hardly deal with domain adaptation, and (ii) qualitatively, where differences between automatic and manual detection processes are studied. **Del Barrio et al.** [5] tackles the disinformation issue from the perspective of understanding how the high-quality European press presents a topic, as it allows understanding of the ecosystem in which disinformation is presented. In particular, the topic investigated is around the Covid-19 vaccination issue. A dedicated dataset of 50k articles is constructed. **Shao et al.** [8] tackles the fake news detection problem by leveraging both single- and multi-modal information from text and images. Classifier outputs are integrated via ensemble learning to increase classification accuracy. Cross-entropy loss is used to maximize the inter-class variations, while the center-loss is used to minimize the intra-class variations. **Tourille et al.** [10] studies the automatic detection of bot-generated Twitter messages, which is in particular a difficult task due to the combination of the strong performance of recent deep language models and the limited length of tweets. They propose a definition of the task, devise two approaches based on pretrained language models, as well as the introduction of a new dataset of generated tweets. **Rakotoson et al.** [7] brings into discussion the scientific community, where scientific facts play an important role. A multi-task approach is introduced for the verification of the scientific questions based on joint reasoning from facts and evidence in research articles. It successfully combines two approaches: (i) an automatic information summarization system and (ii) a Boolean Question Answering approach. The final system is an end-to-end Extractive Question Answering (EQA).

Deepfakes are a very popular means to fuel disinformation. **Baxevanakis et al.** [1] introduce the MeVer deepfake detection service, a web service detecting deep learning manipulations in images and videos which integrates a model ensembling scheme. The design implications and impact of this service are discussed. **Stanciu et al.** [9] investigate the potential of the new family of deep learning methods, namely Capsule Networks, to address the video deepfake detection problem. Their method manages to use a reduced number of parameters while keeping a high accuracy on the predictions. **Coccomini et al.** [3] focus on the challenge of generalizing deepfake detection approaches so that they do not remain tied to one or more specific deepfake generation methods. The audio modality requires different approaches. **Papastergiopoulos et al.** [6] tackle the audio fake detection approaches, creating a new fake audio detection dataset via extending the TIMIT dataset. To this end, various audio representations are investigated, and solutions are devised. **Cuccovillo et al.** [4] address the processing of audio information and introduce techniques for denoising in the context of microphone classification. This would allow audio processing under unfavorable noisy conditions.

ACKNOWLEDGMENTS

This work was supported by the AI4Media H2020 project, partially funded by the European Commission under contract nr 951911.

REFERENCES

[1] Spiros Baxevanakis, Giorgos Kordopatis-Zilos, Panagiotis Galopoulos, Lazaros Apostolidis, Killian Levacher, Ipek Baris Schlicht, Denis Teyssou, Ioannis Kompatsiaris, and Symeon Papadopoulos. 2022. The MeVer DeepFake Detection Service: Lessons Learnt from Developing and Deploying in the Wild. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[2] Jérémie Bogaert, Marie-Catherine de Marneffe, Antonin Descampe, and Francois-Xavier Standaert. 2022. Automatic and Manual Detection of Generated News: Case Study, Limitations and Challenges. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[3] Davide Alessandro Coccomini, Roberto Caldelli, Fabrizio Falchi, Claudio Gennaro, and Giuseppe Amato. 2022. Cross-Forgery Analysis of Vision Transformers and CNNs for Deepfake Image Detection. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[4] Luca Cuccovillo, Antonio Giganti, Paolo Bestagini, Patrick Aichroth, and Stefano Tubaro. 2022. Spectral Denoising for Microphone Classification. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[5] David Alonso del Barrio and Daniel Gatica-Pérez. 2022. How Did Europe's Press Cover Covid-19 Vaccination News? A Five-Country Analysis. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[6] Christoforos Papastergiopoulos, Anastasios Vafeiadis, Ioannis Papadimitriou, Konstantinos Votis, and Dimitrios Tzovaras. 2022. On the Generalizability of Two-dimensional Convolutional Neural Networks for Fake Speech Detection. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[7] Loïc Rakotoson, Charles Letaillieur, Sylvain Massip, and Fréjus Laleye. 2022. Extractive-Boolean Question Answering For Scientific Fact Checking. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[8] Yi Shao, Jiande Sun, Tianlin Zhang, Ye Jiang, Jianhua Ma, and Jing Li. 2022. Fake News Detection Based on Multi-Modal Classifier Ensemble. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[9] Dan-Cristian Stanciu and Bogdan Ionescu. 2022. Uncovering the Strength of Capsule Networks in Deepfake Detection. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*

[10] Julien Tourille, Babacar Sow, and Adrian Popescu. 2022. Automatic Detection of Bot-generated Tweets. In *Proceedings of the 1st ACM International Workshop on Multimedia AI against Disinformation.*