MAD '23 Workshop: Multimedia AI against Disinformation

Luca Cuccovillo, Bogdan Ionescu, Giorgos Kordopatis-Zilos, Symeon Papadopoulos, Adrian Popescu

▶ **To cite this version:**

HAL Id: cea-04489108
https://cea.hal.science/cea-04489108

Submitted on 4 Mar 2024

MAD '23 Workshop: Multimedia AI against Disinformation

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

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

Giorgos Kordopatis-Zilos
Czech Technical University in Prague
Prague, Czech Republic
kordogeo@fel.cvut.cz

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

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

ABSTRACT

With recent advancements in synthetic media manipulation and generation, verifying multimedia content posted online has become increasingly difficult. Additionally, the malicious exploitation of AI technologies by actors to disseminate disinformation on social media, and more generally the Web, at an alarming pace poses significant threats to society and democracy. Therefore, the development of AI-powered tools that facilitate media verification is urgently needed. The MAD '23 workshop aims to bring together individuals working on the wider topic of detecting disinformation in multimedia to exchange their experiences and discuss innovative ideas, attracting people with varying backgrounds and expertise. The research areas of interest include identifying manipulated and synthetic content in multimedia, as well as examining the dissemination of disinformation and its impact on society. The multimedia aspect is very important since content most often contains a mix of modalities and their joint analysis can boost the performance of verification methods.

CCS CONCEPTS

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

KEYWORDS

disinformation detection, artificial intelligence, fact-checking, multimedia forensics, audio forensics, bot detection

1 INTRODUCTION

Modern public communication does not rely anymore solely on mainstream media like newspapers or television, but also takes place over social networks, in real-time, and with live interactions among users. The speedup in the amount of information available, however, also led to an increased amount of misleading content, disinformation and propaganda. Conversely, the fight against disinformation has become even more crucial and demanding, especially for what concerns controversial and crucial topics such as politics, health and religion.

Modern disinformation campaigns are leveraging, among others, off-the-shelf AI-based tools for content creation and modification: hyper-realistic visual, speech, textual and video content have emerged under the collective name of *generative AI*, undermining the perceived credibility of media content. It is, therefore, even more crucial to counter these advances by devising new analysis tools strengthening the capacity of journalists and fact-checkers, but also of the general public, to detect the presence of synthetic and manipulated content, and being robust and trustworthy. Such tools will themselves benefit by recent AI advances in order to offer new capabilities and augment human capacity.

Future multimedia disinformation detection research relies on the combination of different modalities and on the adoption of the latest advances of deep learning approaches and architectures. These raise new challenges and questions that need to be addressed in order to reduce the effects of disinformation campaigns. The workshop, in its second edition, was built on the success of its first edition [3] and welcomed contributions related to different aspects of AI-powered disinformation detection, analysis and mitigation.

2 SCOPE

The call for papers for MAD '23 included the following topics of interest, while leaving open the possibility of contributions of broad relevance to the core theme of the workshop:

- Disinformation detection in multimedia content
- Multimodal verification methods
- Synthetic and manipulated media detection
- Disinformation spread and effects in social media
- Robustness of media verification against adversarial attacks
- Fairness and non-discrimination of disinformation detection
- Temporal and cultural aspects of disinformation
- Open resources, e.g., datasets, software tools
- Multimedia verification systems and applications
- System fusion, ensembling and late fusion techniques
- Benchmarking and evaluation frameworks

3 OBJECTIVE

The second workshop on Multimedia AI against Disinformation (MAD '23) has aimed to bring together individuals who work on the topic of disinformation in multimedia content. The workshop has been conceived as a channel for sharing experiences and working in the field, as well as discussing innovative ideas. Through this interaction and sharing, the workshop has aimed to facilitate advancements in multimodal disinformation detection by promoting communication between individuals who work on different aspects of the issue and with varying areas of expertise, such as AI researchers, social scientists, and fact-checkers.

4 INVITED SPEAKERS

Prof. Roberto Caldelli has been invited to give a talk on "Multimedia Forensics versus disinformation in images and videos: lesson learnt and new challenges". Prof. Caldelli is a Permanent Researcher of the National Inter-University Consortium for Telecommunications (CNIT), coordinates the Multimedia Forensics and Security Laboratory of the CNIT Research Unit at MICC University of Florence, and a Full Professor at Mercatorum University in Rome. His research activity, for over 20 years, has focused on the study of techniques for digital image processing, digital watermarking, cybersecurity, multimedia forensics and adversarial attacks.

Prof. Ioannis Patras has been invited to give a talk on "Controllable image generation and manipulation". Prof. Patras is a professor in Computer Vision and Human Sensing in Queen Mary University of London and the Director of Research Graduate Studies in the school of Electronic Engineering and Computer Science. He has been in the organizing committee of numerous international research conference, is Associate Editor in the journal of Pattern Recognition, Computer Vision and Image Understanding, and Area Chair in all major Computer Vision conferences including, CVPR, ICCV, FG, ICMI, ACII, and BMVC. He has been the primary supervisor for more than 25 PhD students and is a senior member of IEEE.

5 ACCEPTED PAPERS

The workshop accepted 7 submissions tackling various facets of multimedia AI techniques, addressing the analysis of news, social media information, and deepfakes.

Stanciu and Ionescu [8] propose a novel framework to perform data augmentation on deepfake datasets, which combines classic random perturbations of the training content with the application of a random autoencoding networks to the input patches. With the help of the framework, existing state-of-the-art AI models for deepfake detection can be enhanced even further, and generalize to unseen datasets and manipulations. **Dogoulis et al.** [2] address generalization issues of existing models for synthetic image detection by selecting the reference synthetic training content with a novel approach based on a probabilistic quality assessment method. Due to the training content being selected according to its quality, the corresponding detection models are able to perform a harder task than usual, and thus to generalize better when faced with unseen content from different domains. **Papadopoulos et al.** [5] tackle the problem of detecting multimodal disinformation as the misleading combination of an image and its accompanying text, for

which the research community has very scarce data. To address the gap, they assess different choices and methods in generating large amounts of synthetic training examples from pristine data, serving as the basis for the systematic evaluation of multimodal disinformation detection models.

Salvi et al. [6] apply audio folding, i.e., the superposition of multiple audio frames into a single one, to train a new deep-learning model able to distinguish bona fide speech samples from synthetic speech samples. The proposed technique is able to maintain a high detection accuracy despite reducing the processing time by 75%. **Khan and Malik** [4] propose a transformer network able to distinguish bona fide speech samples from synthetic speech samples. The network combines crafted front-end features for spoofing detection and deep features based on an attention mechanism.

Shultz [7] examines how Twitter accounts were used during the Russian-Ukraine conflict to shape English-language conversations about this event. In particular, the author quantifies the correlation between the narratives provided by the Russian government and the ones detected general discussion tweets about the war, focusing on textual content written in English. **Alonso del Barrio and Gatica-Perez** [1] examine how the European press dealt with the no-vax reactions against the Covid-19 vaccine and the disinformation associated with this movement. The analysis demonstrates that the European press actively opposed a variety of hoaxes circulating mainly on social media, and was critical of the anti-vax trend regardless of the political orientation of the news outlet.

ACKNOWLEDGMENTS

REFERENCES
[1] David Alonso del Barrio and Daniel Gatica-Perez. 2023. Examining European Press Coverage of the No-Vax Movement: A NLP framework. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.
[2] Pantelis Dogoulis, Giorgos Kordopatis-Zilos, Symeon Papadopoulos, and Ioannis Kompatsiaris. 2023. Improving Synthetically Generated Images Detection in Cross-Concept Settings. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.
[3] 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*. 688–689.
[4] Awais Khan and Khalid Mahmood Malik. 2023. SpoTNet: A spoofing-aware Transformer Network for Effective Synthetic Speech Detection. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.
[5] Stefanos-Iordanis Papadopoulos, Christos Koutlis, Symeon Papadopoulos, and Panagiotis Petrantonakis. 2023. Synthetic Misinformers: Generating and Combating Multimodal Misinformation. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.
[6] Davide Salvi, Paolo Bestagini, and Stefano Tubaro. 2023. Synthetic Speech Detection through Audio Folding. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.
[7] Benjamin Shultz. 2023. In the Spotlight: The Russian Government's Use of Official Twitter Accounts to Influence Discussions About its War in Ukraine. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.
[8] Dan-Cristian Stanciu and Bogdan Ionescu. 2023. Autoencoder-based Data Augmentation for Deepfake Detection. In *ACM International Workshop on Multimedia AI against Disinformation (MAD '23)*. June 12, 2023, Thessaloniki, Greece.