

Mediate 2023: News Media and Computational Journalism Workshop

Talia Tseriotou¹ Dina Pisarevskaya¹ Elena Kochkina^{1,2} Marya Bazzi^{2,3} Maria Liakata^{1,2,3}
Arkaitz Zubiaga¹

¹ Queen Mary University of London

² The Alan Turing Institute

³ University of Warwick

{t.tseriotou;d.pisarevskaya;m.liakata;a.zubiaga}@qmul.ac.uk

Abstract

The advent of new technologies and platforms has transformed both the production and the consumption of news and information, impacting greatly the social and political landscape internationally. The challenges that result from this unprecedented scale of information diffusion have become apparent to governments, online platforms, and the general public, especially in light of events such as the war in Ukraine, the COVID-19 infodemic and recent political events. During the fourth rendition of MEDIATE, we shed light on the ever-pertinent topic of misinformation research, and studied it via three themes: (1) automated journalism; (2) explainable and multi-modal verification; and (3) content moderation. This workshop brought together media practitioners and academics to discuss research in the context of the three themes, with a particular emphasis on cross-discipline interaction. With three keynotes from both academia and industry, one contributed talk and seven accepted papers, the workshop explored a wide range of topics. While keynote talks covered different aspects of addressing misinformation, they highlighted automation efforts through real world fact-checking infrastructures, multi-modal architectures and different tools including flagging systems beyond big-tech profits to ensure users' safety. The contributed sections presented architectures, datasets, analyses and automation systems towards the detection and remedy of bias, hate speech, misinformation as well as fact-checking in journalism and social media. Overall, the diversity of the presented topics and the overlap across thematic chapters allowed for fruitful discussions and, most importantly, it opened avenues for collaboration and future research.

Introduction

The rise of Web technologies enables most individuals to almost instantaneously disseminate information to a large audience with little-to-no regulation or quality control. This presents benefits to both the media industry and consumers. In particular, digitalisation has reduced the cost of publishing, has built new bridges between media outlets and audiences, and generally has facilitated one's access to information. However, these opportunities come at a price: digital information diffusion tends to amplify misinformation and

polarisation, and makes it hard to distinguish credible information from misleading content. This change has already led multiple disciplines to re-examine the notions of "truth" online. Over the last three years global events, such as the COVID-19 (mis)infodemic, have brought the challenges of online media into the limelight.

As a continuation to the last three editions of the workshop on the same topic at ICWSM, this year's workshop was focused on three interrelated and challenging themes: (1) automated journalism; (2) explainable and multi-modal verification; and (3) content moderation.

Automated journalism: AI-driven automated journalism is becoming mainstream, as more and more news organisations are using it for a variety of different types of stories. According to the "Journalism, Media, and Technology Trends and Predictions" 2022 report from Reuters Institute, AI technologies are fast becoming a core part of a modern news operation at every level – from newsgathering and production right through to distribution¹. Recently, Reuters Institute² also highlighted that extraordinary advances in AI (such as ChatGPT³) provide more opportunities for publishers. But these new technologies will also bring existential and ethical questions, and lead to possible new risks that need to be identified and mitigated.

Explainable and multi-modal verification: Research advancements in automated verification tools aim to address the challenge of effectively processing vast amounts of information towards tackling the scalability constraint of manual verification. The adaptability of these solutions is nevertheless limited due to the lack of human interpretability, transparency and generalisability. Thus, explainable solutions are particularly important towards the enhancement of end-user trust.

Integration of different sources of information and modalities is another important direction in automated information verification. Since posts around rumours often contain both text and content such as images and videos, stimulating increased interest and strong emotions, they promote rumour

¹<https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2022>

²<https://reutersinstitute.politics.ox.ac.uk/journalism-media-and-technology-trends-and-predictions-2023>

³<https://openai.com/blog/chatgpt/>

propagation. It is therefore crucial to adapt rumour verification and fact-checking systems beyond textual information by employing heterogeneous sources and models and exploiting multi-modal signals to enhance current solutions.

Content moderation: News agencies, fact-checking organisations and social media platforms have intensified their efforts to detect and debunk inaccurate information. Identification of hateful and abusive content on social media and in comment sections of news websites remains an open challenge. As a response to online harms, various automated tools are being developed for online media by both academics and companies to assist with human processing and enable faster moderation of vast amounts of content. Human-in-the-loop solutions in particular are being rapidly developed, contributing towards more robust content moderation systems. There are nevertheless still challenges in adopting and generalising such automated solutions in practice.

Towards the examination of these three main themes and their intersection, we simultaneously considered the perspectives and experiences of practitioners and researchers, aiming to identify concrete challenges and opportunities that can be tackled at the cross-section of these two realms.

Keynotes

We had three invited keynote speakers who shared their points of view on the three main themes of the workshop. Each invited speaker often touched on more than one theme throughout their talk.

The first keynote was given by Dr. Guillaume Bouchard, CEO and Co-Founder of CheckStep⁴. The talk proposed methods and approaches for helping platforms to stop the spread of misleading content, including those on a limited budget. The rise of alternative social online platforms and community apps has led to an increasing number of users encountering the same issues as those found on the so-called Big Tech platforms (such as Facebook and Twitter). There is a range of tools available, including AI flagging systems, shared resources, and free APIs, which can be effectively utilized to protect online platforms from online harms, such as misinformation, hate speech, and polarization. In the talk, Dr. Bouchard described challenges and opportunities presented by the online platforms, while also mentioning initiatives, aimed at enabling small and medium-sized organizations to ensure the safety of their users. Best practices for misinformation moderation were suggested: clear community guidelines (including frequent policy updates and compliance with regulations), user reporting enabled (such feedback may be costly to implement but useful on the long run), AI tools usage (basic filters based on large language models, human-in-the-loop solutions etc.), human moderators (including the entire content analysis pipeline, flagging resources and review tools for moderators), clear metrics, and privacy and transparency for users. Specific methods and efforts to tackle multi-modal misinformation (analysing image

and text together, deep fakes detection) were highlighted, using some examples.

The second keynote was given by Prof. Jiebo Luo, a Computer Science Professor at University of Rochester. He outlined applications of misinformation detection in the era of AI-generated content. New challenges arise, such as the difficulty to distinguish between the truth and generative models' hallucinations. This harbours severe risks for many institutions and the very fabric of society by facilitating massive disinformation campaigns and information bubbles. As an example of facts vs. misinformation, Prof. Luo introduced work that focused on understanding the influence of both misinformation and fact-based news on Twitter regarding the COVID-19 vaccine uptake in the U.S. A surprising negative association between the percentage of fact-related users (users who posted tweets with fact-based news rather than fake-news-related tweets) and the vaccination rate was revealed (Lyu, Zheng, and Luo 2022). A possible explanation might be a combination of a larger user-level influence and the negative impact of online social endorsement on vaccination intent. Prof. Luo also discussed AI-generated images and how to detect them. He also described how to incorporate image-related features into a rumour detection framework. Understanding social multimedia content can be combined with information propagation patterns, to detect misinformation. Prof. Luo talked further about research areas of generative AI - for text, image and video generation. Due to the growing prevalence of software that can create AI-generated content, these tools could be used to spread misinformation. It shows the importance of revealing explanatory textual and visual patterns of misinformation.

The third keynote speaker, Dr. Tanu Mitra, is an Assistant Professor at the Information School of the University of Washington. Her keynote concerned human and technological infrastructures of fact-checking, including a system demonstration. Due to the recent increase in scale and diffusion of online misinformation, efforts to develop scalable technological systems for fact-checking online information have also increased. However, such systems are limited in practice because their system design often does not take into account how fact-checking is done in the real world, and they ignore the insights and needs of various stakeholder groups core to the fact-checking process. In the talk, various aspects of this process were revealed: the infrastructures (both human and technological) that support and shape fact-checking work, the primary stakeholders involved in this process, the collaborative effort among them and the associated technological and informational infrastructures (tools, technology and policy that support the stakeholder groups in performing their roles), and finally the key social and technical needs and challenges faced by each stakeholder group. There are, in general, six stakeholder groups: news editors, fact-checkers, copy editors, social media managers, investigators, and advocates. To better understand the infrastructures of fact-checking, results of the interview study with 26 participants from 16 fact-checking teams/organisations from 4 continents were demonstrated. The problem of skepticism towards automation and AI's ability to detect false statements, as well as the need of human-in-the-loop sys-

⁴<https://www.checkstep.com/>

tems and explainable solutions, were also analysed during the talk. Finally, moving to the practical perspective, the preview of the YouCred system was suggested. It was designed to assist fact-checkers with misinformation discovery and credibility assessments on one of the largest video search platforms—YouTube.

Peer-Reviewed Contributions

We had a total of eight contributions: one talk, two short papers and five long papers, spanning across the proposed themes and their cross-section. The papers went through a peer-review process by at least three members of the Program Committee composed of experts in computer science, computational social science, computational linguistics and informatics engineering with domain expertise in misinformation.

Martin Wessel and Timo Spinde from the Media Bias Group gave a talk on “Trends in Automatic Media Bias Detection”. In this talk, Wessel and Spinde presented an overview of the current research developments and trends in media bias detection, with emphasis on data sources and structural gaps. They introduced a Media Bias Taxonomy, for enhanced bias quantification in news articles. They also shed light onto the challenges of this task’s complexity, variability of perception by readers, rarity of datasets, limitations of current models and lack of communication with the public and evaluation, which solidify current research avenues for Media Bias Group. Finally, they introduced 1) their novel datasets, 2) the MBIB project (Wessel et al. 2023) which analyses all existing datasets providing an evaluation platform, 3) MAGPIE, the first large-scale multi-task pre-training approach for media bias detection and 4) a gamified approach for annotations.

The second contribution, entitled “NewsQuote: A Dataset Built on Quote Extraction and Attribution for Expert Recommendation in Fact-Checking” (Zhang et al. 2023), consisted of various approaches towards the task of identifying credible sources for news articles. The authors present the construction process of their novel NewsQuote dataset of 24031 COVID-19 news corpus quote-speaker pairs, while they show between different candidates of source and quotation extraction that a BERT Question-Answering pipeline has the best performance achieving over 98% exact match for source extraction and close to 90% for quote extraction. A main takeaway is that to find credible evidence for news articles, document-centric retrieval is more effective than expert-centric retrieval.

For the third contribution, entitled: “Uncovering Political Hate Speech During Indian Election Campaign: A New Low-Resource Dataset and Baselines” (Jafri et al. 2023), the authors introduced IEHate, a new 11457 manually annotated dataset of Hindi Tweets dataset for Political Hate Speech detection. Given the scarcity of sources they highlight the importance of low-resource Hindi datasets for hate speech detection in political discourse. To further accelerate research on this topic, the dataset is made publicly available. Furthermore, they conducted an analysis and benchmarked the dataset using a range of machine learning and transformer-based algorithms. Although the latter showed

the best performance, the space for performance improvement compared to human evaluation indicates the need for more robust automated systems on hate speech detection during the Indian election campaign and the importance of a mixed approach in hate speech moderation.

In the fourth contribution titled: “Combating the COVID-19 Infodemic: Untrustworthy Tweet Classification using Heterogeneous Graph Transformer” (Ai, Liu, and Hirschberg 2023), the authors presented their contributions on the identification of untrustworthy tweets around COVID-19. They constructed the RTCas-COVID-19 dataset, consisting of 35M tweets with corresponding cascades and weakly labeled 2M of source tweets toward misinformation, providing richer and better social context information corpus (9.5 times higher graph density than other benchmark corpora). Significantly, they contribute towards the under-explored topic of multi-modal verification by modeling linguistic features, user Twitter features and social interactions through community information diffusion. They represent social context as a heterogeneous graph indicating tweeting/retweeting interactions between users, capturing propagation patterns. Towards the identification of trustworthy tweets, they developed RTCS-HGT (a Heterogeneous Graph Transformer enhanced with Retweet Cascade Subgraph Sampling), a SOTA framework for this dataset, achieving 91.8% accuracy and scalability by improving training time up to 93%.

The fifth contribution entitled: “Bias or Diversity? Unraveling Fine-Grained Thematic Discrepancy in U.S. News Headlines” (Pan et al. 2023), examined the thematic discrepancy in US news media, indicative of systematic bias in news reporting. Using the dataset by Lyu et al. (2023) of 1.8M news headlines from nine media outlets, spanning over eight years, they isolated frequently appearing bigrams which were manually categorised with high-agreement into four categories: economics, social issues, foreign affairs and domestic politics. Multiple Correspondence Analysis was performed to measure the thematic discrepancy between the outlets for each n-gram. The analysis revealed different temporal discrepancy patterns across categories while it shed light into specific subtopics that individual outlets paid more attention to, as well as their evolution on an aggregate level. Summarising key takeaways, the authors found temporally decreasing discrepancy on domestic politics, little discrepancy around foreign affairs and economics, but more significant discrepancies on social issues brought about by media bias and attitudinal difference.

The sixth contribution “Leveraging Social Interactions to Detect Misinformation on Social Media” (Fornaciari et al. 2023) was devoted to misinformation threads detection, based on the tweets dataset created during the COVID-19 pandemic. To detect tweet cascades as reliable or unreliable, several multi-input classification models that combined textual features and network features were created: textual BERT model representations + network-sparse-vectors to encode network interactions; BERT embeddings for text + trainable network embeddings; textual BERT representations + Mention2Vec as network representations; and retrofitted BERT + trainable network embeddings. The au-

thors also revealed that harmful threads can contain various topics, and are mostly generated by distinct communities.

The seventh contribution titled “aedFaCT: Scientific Fact-Checking Made Easier via Semi-Automatic Discovery of Relevant Expert Opinions” (Altuncu et al. 2023) introduced aedFaCT. This web browser extension can help professionals and news readers perform scientific fact-checking via the automatic discovery of trusted expert opinions (from news media with high credibility and peer-reviewed scientific publications) relevant to the news of concern via shared keywords. Keywords are extracted automatically, and users are also allowed to add their own keywords. The aedFaCT tool helps minimise manual work (such as using search engines to find evidence) during the fact-checking process. Focusing on evidence retrieval, it leaves the final veracity decision to users.

The eighth contribution “Some Observations on Fact-Checking Work with Implications for Computational Support” (Procter et al. 2023) provided the perspective of fact-checkers and examined daily challenges that fact-checkers face. The study relied on semi-structured interviews with eight members of fact-checking teams from two organisations. The feedback shared by fact-checkers was linked to the different stages of the fact-checking workflow: selecting which claims to check, finding the evidence, reviewing the fact-check, and writing up the fact-check. Application of computational tools in the work of fact-checkers was highlighted. While automation of the whole fact-checking process does not seem to be a feasible goal for fact-checkers, such tools can be used to assist them in their tasks.

Discussion and Future Directions

During the Q&A sessions there were discussions on current and future directions, including joining forces across contributions, some of which are listed below:

- Common user indifference in reading the community guidelines on hateful content and misinformation across platforms unless their activity is affected by them. Therefore intervention by social platforms to enforce these standards is important.
- Discussion around whether the community should invest more resources on deep fake detection or on actor-based detection reached the conclusion that the emphasis should be primarily on actor-based detection (i.e. network analysis), as it is unlikely to achieve a very high accuracy in deep fake detection.
- A learning from language-specific hate speech dataset development that can potentially facilitate extension to other languages and countries is that careful recruiting of annotators that understand the culture and developments and are up to date with the local landscape is key.
- Comparison of methods such as LDA and n-grams for topic extraction on misinformation analysis was held, with the latter being more efficient on short text.
- Tackling misinformation spread through AI generated content requires stepwise understanding of generation methods in order to develop effective solutions.

- Scalability of user-specific systems requires user encoding methods that grow linearly rather than quadratically.
- The primary factors in claim fact-checking prioritisation by experts are: spread, severity and amplification. Detecting claims, the fact-checking of which has been already performed in the past, is of secondary importance.
- The potential of leveraging large generative models such as ChatGPT in fact-checking tasks and especially summarisation was discussed: namely, automated reporting of multiple fact-checked claims to save time.

Workshop Organisation

This workshop was organised as part of the activities of Alan Turing Institute’s special interest group “Media in the Digital Age”⁵. The aim of the interest group is to facilitate collaboration between academics from different communities and practitioners in order to encourage interdisciplinary contributions in the media landscape such as improving the reliability and robustness of the information sphere, technological advancements towards automated journalism and content moderation, scalability and adaptability of solutions and developing a holistic understanding of the state-of-play, allowing a real-world impact. The contributed papers of the workshop are published in the Workshop Proceedings of the International AAI Conference on Web and Social Media⁶.

The Mediate 2023 Workshop was organised by:

- Talia Tseriotou - PhD student at Queen Mary University of London
- Dina Pisarevskaya - PhD student at Queen Mary University of London
- Elena Kochkina - work was done while holding a Post-doctoral Research Assistant position at Queen Mary University of London
- Marya Bazzi - Assistant Professor at The University of Warwick and a Turing Fellow at the Alan Turing Institute
- Maria Liakata - Professor of Natural Language Processing (NLP) at Queen Mary University of London and a UKRI/EPSCRC Turing AI fellow
- Arkaitz Zubiaga - Senior Lecturer (Associate Professor) at Queen Mary University of London

Acknowledgments

This work was supported by a UKRI/EPSCRC Turing AI Fellowship to Maria Liakata (grant EP/V030302/1) and the Alan Turing Institute (grant EP/N510129/1). Talia Tseriotou was supported by the DeepMind PhD Scholarship. Arkaitz Zubiaga was partly supported by the HYBRIDS MSCA Doctoral Network, co-funded by the European Commission and UKRI (grant EP/X036758/1).

⁵<https://www.turing.ac.uk/research/interest-groups/media-digital-age>

⁶<https://workshop-proceedings.icwsm.org/index.php?year=2023>

References

- Ai, L.; Liu, Z.; and Hirschberg, J. 2023. Combating the COVID-19 Infodemic: Untrustworthy Tweet Classification using Heterogeneous Graph Transformer. *ICWSM*.
- Altuncu, E.; Nurse, J. R.; Bagriacik, M.; Kaleba, S.; Yuan, H.; Bonheme, L.; and Li, S. 2023. aedFaCT: Scientific Fact-Checking Made Easier via Semi-Automatic Discovery of Relevant Expert Opinions. *ICWSM*.
- Fornaciari, T.; Luceri, L.; Ferrara, E.; and Hovy, D. 2023. Leveraging Social Interactions to Detect Misinformation on Social Media. *ICWSM*.
- Jafri, F. A.; Siddiqui, M. A.; Thapa, S.; Rauniyar, K.; Naseem, U.; and Razzak, I. 2023. Uncovering Political Hate Speech During Indian Election Campaign: A New Low-Resource Dataset and Baselines. *ICWSM*.
- Lyu, H.; Pan, J.; Wang, Z.; and Luo, J. 2023. Computational Assessment of Hyperpartisanship in News Titles. *arXiv preprint arXiv:2301.06270*.
- Lyu, H.; Zheng, Z.; and Luo, J. 2022. Misinformation versus facts: understanding the influence of news regarding COVID-19 vaccines on vaccine uptake. *Health Data Science, 2022*.
- Pan, J.; Qi, W.; Wang, Z.; Lyu, H.; and Luo, J. 2023. Bias or Diversity? Unraveling Fine-Grained Thematic Discrepancy in U.S. News Headlines. *ICWSM*.
- Procter, R.; Catania, M. A.; He, Y.; Liakata, M.; Zubiaga, A.; Kochkina, E.; and Zhao, R. 2023. Some Observations on Fact-Checking Work with Implications for Computational Support. *ICWSM*.
- Wessel, M.; Horych, T.; Ruas, T.; Aizawa, A.; Gipp, B.; and Spinde, T. 2023. Introducing MBIB—the first Media Bias Identification Benchmark Task and Dataset Collection. *arXiv preprint arXiv:2304.13148*.
- Zhang, W.; Gui, L.; Procter, R.; and He, Y. 2023. NewsQuote: A Dataset Built on Quote Extraction and Attribution for Expert Recommendation in Fact-Checking. *ICWSM*.