Mediate 2021: News Media and Computational Journalism Workshop

Panayiotis Smeros
École Polytechnique Fédérale de Lausanne (EPFL)
panayiotis.smeros@epfl.ch

Jérémie Rappaz
École Polytechnique Fédérale de Lausanne (EPFL)
jeremie.rappaz@epfl.ch

Marya Bazzi
University of Warwick
Alan Turing Institute
mbazzi@turing.ac.uk

Elena Kochkina
Alan Turing Institute
Queen Mary University of London
ekochkina@turing.ac.uk

Maria Liakata
Alan Turing Institute
Queen Mary University of London
mliakata@turing.ac.uk

Karl Aberer
École Polytechnique Fédérale de Lausanne (EPFL)
karl.aberer@epfl.ch

Abstract

With a drastic shift towards digital communication, individuals and organizations are able to almost instantly disseminate information to a large audience with little-to-no regulation, creating both new challenges and opportunities. This digital shift in our mediasphere has caused a profound change in the production and consumption of information, which in turn has strong implications on the social and political landscape. The challenges that result from mass information diffusion have become more visible to the general public in light of recent events such as the COVID-19 infodemic and the US election. In this second rendition of MEDIATE, we continued our focus on the topic of misinformation and examined it via three interrelated lenses: (1) automated methods tackling misinformation; (2) uptake of automated information verification; and (3) ethics, regulation, and governance. In line with the spirit of this workshop series, we brought together media practitioners and academics to discuss these three themes, with particular emphasis on cross-discipline interaction. Our workshop shed light on a variety of perspectives around common themes, opportunities for collaborations and further research, as well as open challenges.

Introduction

Online media today plays an unprecedented role on political, economic, and social scales. The rise of Web technologies enables most individuals to almost instantly disseminate information to a large audience with little-to-no regulation or quality control. This transformation has permanently altered the "information sphere" we live in (Elisa Shearer 2018).

The digital shift presents benefits to both the media industry and the public. In particular, digitalization has reduced the cost of publishing, has built new bridges between media outlets and their audiences, and generally has facilitated access to information. However, these opportunities come at a price: digital information diffusion tends to amplify disinformation and polarization phenomena and makes it hard to distinguish credible information from misleading content (Myllylahti 2018). This change has already led multiple disciplines to re-examine the notions of "truth" online. Over the past year, the COVID-19 infodemic and, more recently, the US election, have further brought the challenges of online media to the attention of the general public.

Key Challenges

We focus on *misinformation* through three interrelated themes: (1) automated and semi-automated methods to counter misinformation; (2) real-world use-cases in which such methods can be employed; and (3) envisioned challenges towards the regulation of such methods. Importantly, we simultaneously considered the perspectives and experiences of practitioners and researchers, attempting to identify concrete challenges and opportunities that can be tackled at the cross-section of these two realms.

Automated methods tackling misinformation. The problem of misinformation spread is one of the most significant challenges of the information age. Social media platforms enable it to spread rapidly, reaching broad audiences before manual verification can be performed. The severe harm that inaccurate information can cause to society in critical situations has led to an increased interest within the scientific community to develop tools that assist with the verification of information from social media. Throughout the workshop, we considered novel methodologies, describing their advantages and challenges in creating accurate systems that would be adopted in practice by journalists and the public. We also highlighted problems in misinformation that require further attention from the academic community.

Uptake of automated information verification. With the rise of high-profile cases of the negative real-world impact of misinformation, news agencies and fact-checking organizations have significantly increased their efforts in debunking false or inaccurate information. Despite such efforts, manual verification is incapable of scaling with the abundance of (mis)information. While researchers have been offering automated solutions, there are a number of challenges to overcome before they have the potential to become widely adopted. Conversations with practitioners have revealed lack of trust in automated solutions for detecting misinformation due to poor *generalisability* to new topics, lack of *interpretability*, and potential *algorithmic bias*. Throughout the workshop, we discussed which automated solutions have been successfully adopted, the challenges in the adoption of solutions addressing misinformation directly, and (partial) solutions to said challenges.

Ethics, regulation, and governance. Human-led moderation is limited by the volume of data that content curators can process. Therefore, automation has the potential to significantly increase the efficiency and scope of content moderation and fact-checking. However, unregulated interventions on social media – typically, content removal – could be regarded as a form of censorship imposed by private companies. This situation could be exacerbated by the introduction of computational methods into moderation processes. Throughout the workshop, we discussed the types of interventions that are the most effective to prevent the spread of false claims, the ethical boundaries of automated content moderation, the limits of automated information regulation, and how one can implement public and private governance on automated content moderation. In addition, the workshop facilitated interdisciplinary discussions around important questions on *digital governance and democracy*.

Keynotes

We had six invited keynote speakers who shared their perspectives on the three main themes of the workshop. We note that while we group key messages from each speaker by theme below, each invited speaker often touched on more than one theme throughout their talk.

Automated Methods tackling Misinformation

Kristina Lerman, Principal Scientist at the USC Information Sciences Institute, talked about evaluating science skepticism and measuring polarization in the US using social media posts (Rao et al. 2020). Her research estimated polarisation across multiple dimensions, including attitudes towards science, politics, and political moderacy using a large-scale COVID-19 Twitter dataset (Chen, Lerman, and Ferrara 2020). Their findings have shown that opinions about COVID-19 are strongly polarized, and polarization dimensions are correlated. For instance, conservatives are more anti-science, while moderates (centrist) are more pro-science. This study found that existing anti-science attitudes (in 2016) created fertile grounds for COVID-19 misinformation and mistrust of experts to spread. Divergent responses to the COVID-19 pandemic in the US showed that partisanship and mistrust of institutions, including science, can increase resistance to COVID-19 mitigation measures and vaccine hesitancy.

Chris Bregler Director / Principal Scientist at Google AI gave a talk titled "Context is everything: On Manipulated Media, Context Retargeting, and Misinformation Mitigation". The key point of the talk was to raise awareness of the danger posed by so-called *cheapfakes*[1]. While deepfakes are indeed dangerous and certainly enable various forms of abuse, cheapfakes, are more prevalent and require more attention than they currently receive from academics. Cheapfakes are also more difficult to detect as they are more general. While more researchers have moved into cheapfakes over the past year, an important aspect of them that still has not garnered enough attention is the change of context (e.g., change of caption, time, location) in original material (Aneja, Bregler, and Nießner 2021). For this purpose, a new challenge to encourage research in out-of-context detection has been announced.[2]

Mevan Babakar, COO at FullFact[3] and Board Member of Democracy Club & International Fact-Checking Network, talked about how fact-checking is performed at FullFact and how it is automated (Babakar and Moy 2016). Three tasks are being automated at FullFact: (1) claim detection, i.e., identifying factual statements that can be checked from the daily stream of sentences from UK media; (2) claim matching, i.e., identifying repetitions of claims, and (3) robochecking, i.e., checking a claim in real-time against primary sources. From a methodology perspective, both claim matching and claim detection solutions currently in use at FullFact rely on BERT-based models. Specifically, claim matching involves an ensemble of Sentence-BERT models for semantic similarity, topic detection, and entity extraction. Mevan pointed out that claim matching and claim detection are synergistic tasks.

Uptake of Automated Information Verification

Mevan Babakar pointed out that FullFact has experienced an impressive increase in their ability to fact check information since their adoption of automated claim detection, which boosted the number of detected claims by 1000x. However, she highlighted the need for accountability at every level of information dissemination and the fact-checking process. She also emphasized the importance that fact-checking organizations retain independence from governmental and other private organizations. Furthermore, Mevan mentioned a range of crucial and still open questions about the fact-checking process that will inform the future direction of discussions and research in this field, such as how to prioritize the most valuable claims, how third parties are using the data, and how platform-wide changes affect long term behaviors and attitudes.

Rasmus Kleis Nielsen, Professor at the University of Oxford and Director of the Reuters Institute gave a talk on "'News you don't believe': Public perspectives on f*ke news and misinformation and what they can tell us about automated and regulatory responses". Rasmus noted that workable and robust mitigants to problems in online media such as misinformation lie at the intersection of (1) what research can validate and improve, (2) what policy can endorse, and (3) what the public needs or understands the problem to be. Studies from the Reuters Institute for the Study of Journalism[4] surveyed individuals from a number of countries and contains statistics and qualitative insights into different facets of this question (N. Newman et al. 2020; Nielsen and Graves 2017). One important finding of these studies is that people's conception of fake news is much broader than the commonly studied interpretation of "false

[1] an audiovisual manipulation created with cheaper, more accessible software, e.g., Photoshop, lookalikes, re-contextualizing footage, speeding, or slowing (Paris and Donovan 2019)

[2] https://2021.acmmmsys.org/cheapfake_challenge.php

[3] https://fullfact.org

[4] https://www.digitalnewsreport.org

information"; it includes satire, poor journalism, and certain types of advertising. Another finding of these studies is that there is significant public exposure to "poor journalism" (e.g., factual mistakes, misleading headlines, clickbaits). For policy-level attempts at tackling misinformation, it is crucial to be evidence-based but also to account for the public's perception and understanding of misinformation.

Ethics, Regulation, and Governance

David Leslie, Ethics Theme Lead at the Alan Turing Institute, gave a talk on "Governing Critical Digital Infrastructure as a Global Public Utility" (Leslie 2020; Leslie 2019). He drew from existing work and ideas and highlighted five key practical goals that regulators should think about in safeguarding digital governance: (1) prohibiting of targeted advertising; (2) securing common carriage, equity, fair pricing, and non-discrimination; (3) setting in place structural regulations that remove incentives to and organization enablers of predatory behavior; (4) instituting mechanisms of democratic governance and community ownership; and (5) building a global regulatory capacity to manage the global character of critical digital infrastructures. David also mentioned the importance of researcher awareness for mitigating potential misuse of automated systems and the ongoing work in this space by the Alan Turing Institute.

Ramsha Jahangir, Journalist at Coda Story and Scholar at Erasmus Mundus Journalism, gave a talk on "Content regulation in the digital age". Ramsha highlighted the tension between government regulations, platform interventions, and internship freedoms, e.g., governments can ban media platforms, and media platforms can de-platform individuals, control misinformation labeling, and ban political advertisements. All these actions have repercussions on internet freedoms. While the tension between government, platforms, and internet freedoms is present on a global scale, it is crucial to remember that it can manifest with very different degrees of citizen censorship in different parts of the world (Jahangir 2020). The way in which these three entities should interact in the years to come is a big open question with profound ramifications.

Contributed Papers

We had two contributions on automated methods tackling misinformation and three contributions on the uptake of automation, discussing potential solutions that can be implemented by social media platforms to combat the spread of misinformation.

Automated Methods tackling Misinformation

In the first contribution on this topic (Gruppi et al. 2021), the authors study the utilization of tweets by news sources of different credibility levels. Specifically, the authors show the differences in tweets embedded by reliable and unreliable sources regarding quantity and quality, the topics they cover, and the individuals they cite. The main takeaway of this study is that unreliable sources use significantly more Twitter-based content than reliable sources. Furthermore, 41% of the users cited by reliable sources were accounts verified by Twitter, against 14% cited by unreliable sources.

In the second contribution (Gruppi, Horne, and Adalı 2021), the authors propose a novel news veracity detection model. In particular, the authors show that content sharing behaviors, formulated as networks, represent signals of reliability. They investigate the interplay between network and text features in a predictive task and show that incorporating content sharing features leads to performance gain and makes the model more robust to concept drifts.

Uptake of Automated Information Verification

Mohsen Mosleh gave a talk on "Reducing Inaccurate Information on Social Media". In this talk, Mohsen described several cognitive reflection experiments deployed on Twitter to study the prevalence of fundamentally low-quality information. Specifically, the experiments focused on two research questions, namely, who you follow and what you share. The main takeaway of these experiments was that people who engage in less cognitive reflection are more likely to consume and share low-quality content.

In another contribution on the same topic (Sumpter and Neal 2021), the authors conducted a cross-sectional survey in which 204 survey respondents rated the credibility of four news articles, each randomly assigned a credibility warning (i.e., an assessment of the article's credibility determined by either an AI agent or human journalist, or no assessment at all). The authors found that AI warnings are as successful, if not more so, than warnings provided by a journalist, at influencing participants' assessments of a news article's credibility (regardless of the warning's accuracy). Furthermore, language sentiment may influence the degree to which a user perceives and believes such warnings.

In the last contribution (Gausen, Luk, and Guo 2021), the authors developed an agent-based model of the spread of information on social media networks with four agent types (susceptible, believer, denier, and cured). The main takeaway of this work was that agent-based modeling could be a useful tool for policy-makers to evaluate misinformation countermeasures at scale before implementing them on a social media platform.

Discussion and Future Directions

Several opportunities and open questions were highlighted during the workshop. We mention a few below:

- Discussions throughout the workshop have highlighted that communication of uncertainty and explanations of model decisions will play a key role in unlocking public trust in automated fact-checking.
- The need for accountability at every level of information dissemination and fact-checking process, as well as the importance of the independence of fact-checking organizations, were also highlighted during the workshop.
- An array of important open questions to enhance and enable trustworthy fact-checking decisions and interventions were presented, such as how to prioritize the most valuable claims, what is an acceptable margin of error in

automated solutions, how third parties are using the fact-checked information, and how platform-wide changes affect long-term behaviors and attitudes.

- In light of the current pandemic, when ongoing research unravels new facts at a high pace in real-time, there is a rising concern on how automated verification tools will tackle time-sensitive claims. This highlights the need for further research in this direction.
- While deepfakes are dangerous and leave room for serious forms of abuse, cheapfakes and, in particular, out-of-context cheapfakes are far more prevalent "in the wild". Further academic effort and attention is needed to automate the detection of out-of-context cheapfakes. In general, researchers should not ignore seemingly easier but yet unresolved problems, and take into account their prevalence in the real world.
- Nudging users to consider the accuracy of the information affects how users consume and share content. Measuring and comparing the causal effect of various intervention methods, such as misinformation labeling or nudging, on the way in which individuals interact with information, remains an important open question.
- Credible and robust mitigants for misinformation need to lie at the intersection of (1) what research can validate and improve, (2) what policy can endorse, and (3) what the public perception of the problem is.
- Measures taken by governments and platforms differ significantly across countries. Therefore, it becomes apparent that efforts to tackle misinformation need to account for the circumstances and needs of different countries, as well as the scale of the problem.
- Misinformation is a multifaceted problem where different entities play a key role: government, policy makers, platforms, publishers. The way in which the incentives of these different components should be balanced against each other and against the freedoms of citizens is an open and ongoing discussion with deep ramifications.

Workshop Organization

This workshop was organised as part of the efforts of Alan Turing Institute's special interest group **"Media in the Digital Age"**.[5] The organizers of the interest group have a multifaceted expertise in research areas related to news media such as rumor detection and verification (Kochkina, Liakata, and Zubiaga 2018; Kochkina and Liakata 2020; Gorrell et al. 2019), network science (Bazzi et al. 2020; Bazzi et al. 2016), selection bias (Rappaz, Bourgeois, and Aberer 2019; Bourgeois, Rappaz, and Aberer 2018) and scientific misinformation (Smeros, Castillo, and Aberer 2019; Romanou et al. 2020). The purpose of the interest group is to bring together and facilitate concrete collaborations between academics and practitioners (e.g., journalists, fact-checkers, and platforms) in order to explore important topics such as the adoption of technology in the media sphere, misinformation, content moderation, and personalization, interplay between key players of the mediasphere, the future of digital media.

Panayiotis Smeros, Jérémie Rappaz, Marya Bazzi, Elena Kochkina, Maria Liakata, and Karl Aberer served as co-chairs of this workshop. The contributed papers of the workshop are published in the Workshop Proceedings of the International AAAI Conference on Web and Social Media. [6]

[5] https://www.turing.ac.uk/research/interest-groups/media-digital-age, https://digitalmediasig.github.io

[6] http://workshop-proceedings.icwsm.org

Acknowledgements

Work by Maria Liakata and Elena Kochkina was supported by a UKRI/EPSRC grant (EP/V048597/1) to Profs Yulan He and Maria Liakata as well as project funding from the Alan Turing Institute, grant EP/N510129/1. Marya Bazzi was supported by the Alan Turing Institute under the EPSRC Grant No. EP/N510129/1.

References

[Aneja, Bregler, and Nießner 2021] Aneja, S.; Bregler, C.; and Nießner, M. 2021. COSMOS: Catching Out-of-Context Misinformation with Self-Supervised Learning. *arXiv preprint arXiv:2101.06278*.

[Babakar and Moy 2016] Babakar, M., and Moy, W. 2016. The state of automated factchecking: How to make factchecking dramatically more effective with technology we have now. *Full Fact* 28.

[Bazzi et al. 2016] Bazzi, M.; Porter, M. A.; Williams, S.; McDonald, M.; Fenn, D. J.; and Howison, S. D. 2016. Community detection in temporal multilayer networks, with an application to correlation networks. *SIAM Multiscale Model. Simul.* 14(1):1–41.

[Bazzi et al. 2020] Bazzi, M.; Jeub, L. G. S.; Arenas, A.; Howison, S. D.; and Porter, M. A. 2020. A framework for the construction of generative models for mesoscale structure in multilayer networks. *Physical Review Research* 2:023100.

[Bourgeois, Rappaz, and Aberer 2018] Bourgeois, D.; Rappaz, J.; and Aberer, K. 2018. Selection bias in news coverage: learning it, fighting it. In *Companion Proceedings of the The Web Conference 2018*, 535–543.

[Chen, Lerman, and Ferrara 2020] Chen, E.; Lerman, K.; and Ferrara, E. 2020. Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health and Surveillance* 6(2):e19273.

[Elisa Shearer 2018] Elisa Shearer. 2018. Social media outpaces print newspapers in the U.S. as a news source. *Pew Research Center*.

[Gausen, Luk, and Guo 2021] Gausen, A.; Luk, W.; and Guo, C. 2021. Can we stop fake news? using agent-based modelling to evaluate countermeasures for misinformation on social media. ICWSM.

[Gorrell et al. 2019] Gorrell, G.; Kochkina, E.; Liakata, M.; Aker, A.; Zubiaga, A.; Bontcheva, K.; and Derczynski, L.

2019. Semeval-2019 task 7: Rumoureval, determining rumour veracity and support for rumours. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, 845–854.

[Gruppi et al. 2021] Gruppi, M.; Adalı, S.; Salemi, M.; and Horne, B. D. 2021. From tweeting about news to creating news around tweets: Characterizing tweets embedded in news articles. ICWSM.

[Gruppi, Horne, and Adalı 2021] Gruppi, M.; Horne, B. D.; and Adalı, S. 2021. Tell me who your friends are: Using content sharing behavior for news source veracity detection. ICWSM.

[Jahangir 2020] Jahangir, R. 2020. Pakistan's Tinder ban signals coming showdowns with YouTube and Twitter. https://www.codastory.com/authoritarian-tech/pakistans-digital-crackdown/.

[Kochkina and Liakata 2020] Kochkina, E., and Liakata, M. 2020. Estimating predictive uncertainty for rumour verification models. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6964–6981. Online: Association for Computational Linguistics.

[Kochkina, Liakata, and Zubiaga 2018] Kochkina, E.; Liakata, M.; and Zubiaga, A. 2018. All-in-one: Multi-task learning for rumour verification. In *Proceedings of the 27th International Conference on Computational Linguistics*, 3402–3413.

[Leslie 2019] Leslie, D. 2019. Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of ai systems in the public sector. *Available at SSRN 3403301*.

[Leslie 2020] Leslie, D. 2020. Tackling covid-19 through responsible ai innovation: Five steps in the right direction. *Harvard Data Science Review (2020)*.

[Myllylahti 2018] Myllylahti, M. 2018. An attention economy trap? an empirical investigation into four news companies' facebook traffic and social media revenue. *Journal of Media Business Studies* 15(4):237–253.

[N. Newman et al. 2020] N. Newman, R. F.; Schulz, A.; Andi, S.; and Nielsen, R. K. 2020. Digital news report 2020. *Reuters Institute for the Study of Journalism*.

[Nielsen and Graves 2017] Nielsen, R. K., and Graves, L. 2017. News you don't believe: Audience perspectives on fake news. *Reuters Institute for the Study of Journalism*.

[Paris and Donovan 2019] Paris, B., and Donovan, J. 2019. Deepfakes and cheap fakes: The manipulation of audio and visual evidence. *Data & Society*.

[Rao et al. 2020] Rao, A.; Morstatter, F.; Hu, M.; Chen, E.; Burghardt, K.; Ferrara, E.; and Lerman, K. 2020. Political partisanship and anti-science attitudes in online discussions about covid-19. *arXiv preprint arXiv:2011.08498*.

[Rappaz, Bourgeois, and Aberer 2019] Rappaz, J.; Bourgeois, D.; and Aberer, K. 2019. A dynamic embedding model of the media landscape. In *The World Wide Web Conference*, 1544–1554.

[Romanou et al. 2020] Romanou, A.; Smeros, P.; Castillo, C.; and Aberer, K. 2020. Scilens news platform: A system for real-time evaluation of news articles. *Proc. VLDB Endow.* 13(12):2969–2972.

[Smeros, Castillo, and Aberer 2019] Smeros, P.; Castillo, C.; and Aberer, K. 2019. Scilens: Evaluating the quality of scientific news articles using social media and scientific literature indicators. In Liu, L.; White, R. W.; Mantrach, A.; Silvestri, F.; McAuley, J. J.; Baeza-Yates, R.; and Zia, L., eds., *The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019*, 1747–1758. ACM.

[Sumpter and Neal 2021] Sumpter, M., and Neal, T. 2021. User perceptions of article credibility warnings: Towards understanding the influence of journalists and ai agents. ICWSM.