The 1^{st} Workshop on Misinformation Detection in the Era of LLMs (MisD 2025)

Zhiwei Liu[1] **Christine de Kock**[2] **Peter Knight**[1]
Eduard Hovy[2,3] **Sophia Ananiadou**[1]
[1]The University of Manchester, Manchester, UK
[2]The University of Melbourne, Australia
[3]Carnegie Mellon University, USA
{zhiwei.liu, peter.knight, sophia.ananiadou}@manchester.ac.uk
{christine.dekock,eduard.hovy}@unimelb.edu.au

Abstract

The rise of the internet and social media has made it easy to spread misinformation, resulting in a large impact on society. Furthermore, while the emergence of large language models (LLMs) brings hope to detection efforts, it also facilitates the generation of false information, making such content increasingly indistinguishable from reality. Therefore, there is an increasing urgency for further research on misinformation detection in the era of LLMs. In this workshop, we explore the potential of LLMs to address complex mis/disinformation detection issues and facilitate discussions on the current state and future directions in misinformation detection and understanding, and drive the development of comprehensive frameworks that address the multifaceted nature of misinformation detection challenges.

Introduction

With the rise of social media such as X, Facebook, and Weibo, an increasing number of people are browsing information online. The latest statistics show that active social media user identities have passed the 5 billion mark, with the latest user figure equivalent to 63.9 percent of the world's population[1]. However, due to lagging of online regulation, the internet is flooded with a large amount of misinformation, including fake news, rumors, and conspiracy theories (Scheufele and Krause 2019) (latest examples like Facebook and Instagram get rid of fact-checkers[2]). Such false pieces of information, or misleading combinations of factual information to support unwarranted conclusions, lead people to believe in unreal content, drive public opinion, and pose serious harm to society, the economy, and politics (Petratos 2021). Furthermore, the latest advances in Artificial Intelligence (AI) and large language models (LLMs) such as ChatGPT and GPT-4 (Achiam et al. 2023) have made it easier to generate seemingly persuasive false information (Barman, Guo, and Conlan 2024). Therefore, there is an urgent global need for methods that can effectively detect erroneous and misleading information.

[1]https://datareportal.com/reports/digital-2025-global-overview-report

[2]https://www.bbc.co.uk/news/articles/cly74mpy8klo

LLMs have significantly advanced the field of misinformation detection by enhancing the efficiency and accuracy of predictive models (Huang et al. 2025b; Liu et al. 2024; Chen and Shu 2024). However, using LLMs for misinformation detection still faces many challenges, including scalability, bias, contextual understanding, interpretability, and adaptability to new types of fake content (Augenstein et al. 2024). Also, they can be used to generate convincingly fake content on a large scale. Furthermore, given known issues with hallucination in LLMs, there is a need for consideration of how much automation is feasible (Huang et al. 2025a).

Although there are currently some related workshops and shared tasks, such as CheckThat! Lab (Barrón-Cedeño et al. 2024) (which focuses on the information verification pipeline), and FEVER (Schlichtkrull et al. 2024) (a venue for work in verifiable knowledge extraction), these initiatives primarily focus on fact-checking methods, while areas such as misleading conclusions from factually accurate content, propaganda, practical issues with content moderation, and policies remain less explored. Additionally, recent efforts such as the Financial Misinformation Detection Challenge (Liu et al. 2025) have begun to address domain-specific challenges, emphasizing the importance of misinformation in high-stakes areas like finance. These contributions also suggest the need for a more comprehensive and interdisciplinary approach to address misinformation more broadly.

This workshop explore the potential of LLMs to address such complex mis/disinformation detection issues and its implications for content moderation systems. The workshop will facilitate discussions on the current state and future directions of NLP techniques in misinformation detection and understanding, and drive the development of comprehensive frameworks that address the multifaceted nature of misinformation detection challenges.

TOPICS AND THEMES

- Methodology – Applying LLMs to identify fake news, rumors, or conspiracy theories.
  - Fact checking - Determining the 'truth' of claims against given background references.
  - Multi-modal/multi-lingual misinformation detection - Leverage different modalities/languages and combi-

nations thereof to tackle online multimodal offensive content.
  - Cross-domain misinformation detection - identify misinformation collected from health, education, finance, politics, technologies, etc.
  - Stance detection - Identifying topics and sentiment/emotions.
  - Network analysis - Analyzing social networks, dissemination methods, etc., of misinformation.
- Implication - Developing methods to identify misleading reasoning that uses true facts but leads to unwarranted conclusions.
- Interpretability - Providing explanations when detecting misinformation or fact-checking.
- Feature analysis - Analyzing the impact of different features for misinformation detection, such as emotion, style, stance, etc.
- Hallucination mitigation and evaluation in LLMs.
- Data source and benchmark - Contributions of new datasets and benchmarks and analysis of the misinformation generated by LLMs.
- Fairness of LLM moderation: Existing work has shown that LLMs exhibit systematic biases against different demographics (e.g. religion, age, or other cultural characteristics). To what extent does this impact misinformation detection?

Accepted Submissions and Format

This workshop focuses on the detection of misinformation in the era of large language models. The workshop attracted a diverse and interdisciplinary audience, including researchers from computer science and the social sciences, as well as professionals and academics from both industry and academia.

All submitted papers underwent a rigorous peer-review process. Each submission was reviewed by at least three independent reviewers and evaluated based on originality, relevance, and technical soundness. Following thorough deliberations by the program committee, seven high-quality papers were accepted for inclusion in the workshop proceedings.

- Checklist-Prompted Feature Extraction for Interpretable and Robust Fact Check Worthiness Prediction.
  Yuka Teramoto, Takahiro Komamizu, Mitsunori Matsushita, Kenji Hatano
- Multi-agent Systems for Misinformation Lifecycle: Detection, Correction and Source Identification
  Aditya Gautam
- From Millions of Tweets to Actionable Insights: Leveraging LLMs for User Profiling
  Vahid Rahimzadeh, Ali Hamzehpour, Azadeh Shakery, Masoud Asadpour
- Towards a Multi-modal Multi-Label Election-Context Repository for Classifying Misinformation
  Prerana Khatiwada, Qile Wang, Kenneth E. Barner, Matthew Louis Mauriello
- SegSub: Evaluating Robustness to Knowledge Conflicts and Hallucinations in Vision-Language Models
  Peter Carragher, Nikitha Rao, Abhinand Jha, R Raghav, Kathleen M. Carley
- An Analytical Emotion Framework of Rumour Threads on Social Media
  Rui Xing, Boyang Sun, Kun Zhang, Preslav Nakov, Timothy Baldwin, Jey Han Lau
- Misleading through Inconsistency: A Benchmark for Political Inconsistencies Detection
  Nursulu Sagimbayeva, Ruveyda Betul Bahceci, Ingmar Weber

This workshop was a one-day workshop. Our agenda includes two keynote invited talks, 9 oral presentations (including the accepted papers and invited presentations), and two discussion sessions highlighting research opportunities, pioneering solutions, systems, success trajectories, and visionary future directions in the coming years.

Invited Speakers

We are honored to have invited two distinguished experts to contribute to the workshop: Dr. Johan Farkas, representing the field of social science, and Dr. Dustin Wright, from the domain of computer science. Their diverse disciplinary backgrounds and deep expertise provide valuable perspectives that enrich the workshop's interdisciplinary discourse on misinformation detection in the era of large language models.

- **Johan Farkas** is an Assistant Professor in Media Studies at the University of Copenhagen. His research explores the intersection of digital media, journalism, disinformation, and democracy. Farkas is the author of Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehood (written with Jannick Schou), which presents a critical examination of discourses around fake news and the post-truth era.
  - **Keynote Title**: The Democratic Deficit of Misinformation Solutions
  - **Abstract**: Contemporary solutions to misinformation and disinformation—from automated fact-checking to content moderation, algorithmic nudging to policy interventions—are often framed by scholars, journalists, and politicians as efforts to protect democracy. Yet these interventions often rest on a narrow vision of what democracy is and ought to be. In this keynote, Johan Farkas interrogates the underlying normative assumptions of misinformation responses across journalism, academia, and politics. Farkas argues that proposed and implemented solutions tend to reproduce a one-sided, ahistorical understanding; one that equates democracy with rationality and expert authority while sidelining key democratic values such as egalitarianism, participation, and popular sovereignty.
    Farkas shows how prevailing approaches to mis- and disinformation often implicitly legitimise an increasingly authoritarian model of governance, in which

public participation and popular rule are rendered secondary to decision-making by politico-economic elites. This contributes to already-existing democratic deficits across liberal democracies as well as ongoing processes of hollowing out of democratic participation.

The lecture builds on the book Post-Truth, Fake News and Democracy: Mapping the Politics of Falsehoods (co-authored with Jannick Schou, published by Routledge in a second edition in 2023)

- **Dustin Wright** is a Danish Data Science Academy postdoctoral fellow at the University of Copenhagen. His research is driven by a central mission: to make the world's knowledge reliable and accessible. This mission spans real-world knowledge, scientific knowledge, and diverse forms of human knowledge. He draws on methods from machine learning and natural language processing to achieve this goal. His work on reliability focuses on ensuring the factuality and faithfulness of various types of knowledge, as well as promoting the safety and interpretability of machine learning systems. On the accessibility side, he aims to develop machine learning models that process knowledge efficiently and sustainably.
  - **Keynote Title**: The many faces of science misinformation
  - **Abstract**: Science allows us to discover new knowledge about out world. Sharing this knowledge with the public, for example through science communication, is critical to build an informed society and establish trust in the institution of science. Scientific misinformation has the opposite effect, from degrading the public trust in science to endangering people's health. In this talk, Dustin Wright presents his work which aims to use natural language processing (NLP) to understand and combat it. Dustin discusses scientific misinformation in its multiple forms, including categorical falsehoods, exaggerations, and sensationalism, and the different methods needed to detect them. Finally, He discusses the potential for NLP to help us better understand and improve the science communication process through modeling the public perception of science news.

List of Program Committee Members

Our Program Committee members consist of distinguished reviewers from both academia and industry. The list is as follows:

- Calvin Cheng - University of Oxford
- Aditya Gautam - Meta
- Mohsinul Kabir - University of Manchester
- Polydoros Giannouris - University of Manchester
- Aman Rangapur - Allen Institute for Artificial Intelligence
- Zukang Yang - University of California, Berkeley
- Yupeng Cao - Stevens Institute of Technology
- Rynaa Grover - Google
- Dipanwita Guhathakurta - Georgia Institute of Technology

Workshop Organizers

The workshop was organized by researchers from the University of Manchester and the University of Melbourne.

- **Eduard Hovy** is the Executive Director of Melbourne Connect at the University of Melbourne. Prof. Hovy holds adjunct professorships in Carnegie Mellon University (CMU)'s Machine Learning Department and at the University of Southern California (Los Angeles). He is one of the initial 17 Fellows of the ACL and is also a Fellow of the AAAI. Prof. Hovy's research focuses on computational semantics of language and addresses various areas in Natural Language Processing and Data Analytics, including in-depth machine reading of text, information extraction, automated text summarisation, question answering, the semi-automated construction of large lexicons and ontologies, and machine translation. Prof. Hovy is the author or co-editor of eight books and over 450 technical articles. He has an h-Index of 109 (Google Scholar) with more than 68,000 citations.
- **Sophia Ananiadou** is Professor of Computer Science at the University of Manchester. She is the Director of the UK National Centre for Text Mining and holds roles such as Deputy Director of the Institute of Data Science and AI (Manchester), ELLIS fellow, Distinguished Research Fellow at the AI Research Centre (AIST Japan), and Lead Researcher at the Archimedes Research Centre, Athens, Greece. Her research focus is on leveraging NLP to understand and utilize language knowledge, in special domains such as biomedicine, health and finance for the tasks of information extraction, summarisation, emotion and misinformation detection. Her h-Index is 75 (Google Scholar) with 23,000 citations. She has organized numerous workshops and shared tasks (e.g. BioNLP, CL4Health) such as ACL, EMNLP, NeurIPS, IJCAI, LREC and COLING. She is one of the founders of SIGBioMed (ACL) and SIG-FinTech (ACL).
- **Christine de Kock** is a Doreen Thomas fellow and lecturer in the School of Computing and Information Systems at the University of Melbourne. Her research interests include online communities, extremism and disinformation. She is a convenor on the Hallmark Research Initiative on Fighting Harmful Online Information, a 3-year interdisciplinary initiative combining experts from psychology, media studies, philosophy and computer science to combat modern challenges around online harms. She has publications in top-tier conferences such as ACL, EMNLP, EACL. She is an organiser on the Workshop for Online Abuse and Harms (ACL 2025), the Workshop for Protecting Women Online (WWW 2025), and two SemEval shared tasks focused on African languages (Task 1 2024 and Task 11 2025).
- **Peter Knight** is professor of American Studies at the University of Manchester, and has also held visiting fellowships at New York University, Harvard, the Smithsonian, Leiden University and the Netherlands Institute

for Advanced Studies. His research is on conspiracy theories, and the economic humanities. He is the author and editor of a dozen books, including Conspiracy Culture: From the Kennedy Assassination to "The X-Files" (2000), Conspiracy Theories in the Time of Covid-19 (Routledge, 2022) and The Routledge Handbook of Conspiracy Theories (2020). He directed the EU-funded Comparative Analysis of Conspiracy Theories network; Infodemic, a project funded by the UKRI on conspiracy theories and the pandemic; Everything Is Connected, an AHRC-funded project on conspiracy theories in the age of the internet; and is a co-investigator on the EU-funded REDACT and TACT projects.

- **Zhiwei Liu** is a PhD candidate at the Department of Computer Science at the University of Manchester. He focuses on the technical applications and discoveries of LLMs, primarily applied in misinformation detection and sentiment analysis. He has already published related papers in top journals or conferences, such as ACL, SIGKDD, NeurIPS, WWW, ECAI, and Information Fusion. He is the co-organizer of FinNLPAgentScen@IJCAI-2024 workshop and COLING2025 Financial Misinformation Detection Challenge, highlighting his organizational and coordination skills as well as his determination to drive the development of NLP and the field of misinformation detection.

Acknowledgments

This work is supported by the Manchester-Melbourne-Toronto Research Fund and the Centre for Digital Trust and Society at the University of Manchester.

References

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Augenstein, I.; Baldwin, T.; Cha, M.; Chakraborty, T.; Ciampaglia, G. L.; Corney, D.; DiResta, R.; Ferrara, E.; Hale, S.; Halevy, A.; et al. 2024. Factuality challenges in the era of large language models and opportunities for fact-checking. *Nature Machine Intelligence*, 6(8): 852–863.

Barman, D.; Guo, Z.; and Conlan, O. 2024. The dark side of language models: Exploring the potential of llms in multimedia disinformation generation and dissemination. *Machine Learning with Applications*, 100545.

Barrón-Cedeño, A.; Alam, F.; Chakraborty, T.; Elsayed, T.; Nakov, P.; Przybyła, P.; Struß, J. M.; Haouari, F.; Hasanain, M.; Ruggeri, F.; Song, X.; and Suwaileh, R. 2024. The CLEF-2024 CheckThat! Lab: Check-Worthiness, Subjectivity, Persuasion, Roles, Authorities, and Adversarial Robustness. In Goharian, N.; Tonellotto, N.; He, Y.; Lipani, A.; McDonald, G.; Macdonald, C.; and Ounis, I., eds., *Advances in Information Retrieval*, 449–458. Cham: Springer Nature Switzerland. ISBN 978-3-031-56069-9.

Chen, C.; and Shu, K. 2024. Combating misinformation in the age of llms: Opportunities and challenges. *AI Magazine*, 45(3): 354–368.

Huang, L.; Yu, W.; Ma, W.; Zhong, W.; Feng, Z.; Wang, H.; Chen, Q.; Peng, W.; Feng, X.; Qin, B.; et al. 2025a. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2): 1–55.

Huang, T.; Yi, J.; Yu, P.; and Xu, X. 2025b. Unmasking Digital Falsehoods: A Comparative Analysis of LLM-Based Misinformation Detection Strategies. *arXiv preprint arXiv:2503.00724*.

Liu, Z.; Wang, K.; Bao, Z.; Zhang, X.; Dong, J.; Yang, K.; Kabir, M.; Giannouris, P.; Xing, R.; Park, S.; Kim, J.; Li, D.; Xie, Q.; and Ananiadou, S. 2025. FinNLP-FNP-LLMFinLegal-2025 Shared Task: Financial Misinformation Detection Challenge Task. In Chen, C.-C.; Moreno-Sandoval, A.; Huang, J.; Xie, Q.; Ananiadou, S.; and Chen, H.-H., eds., *Proceedings of the Joint Workshop of the 9th Financial Technology and Natural Language Processing (FinNLP), the 6th Financial Narrative Processing (FNP), and the 1st Workshop on Large Language Models for Finance and Legal (LLMFinLegal)*, 271–276. Abu Dhabi, UAE: Association for Computational Linguistics.

Liu, Z.; Yang, K.; Xie, Q.; de Kock, C.; Ananiadou, S.; and Hovy, E. 2024. RAEmoLLM: Retrieval Augmented LLMs for Cross-Domain Misinformation Detection Using In-Context Learning based on Emotional Information. *arXiv preprint arXiv:2406.11093*.

Petratos, P. N. 2021. Misinformation, disinformation, and fake news: Cyber risks to business. *Business Horizons*, 64(6): 763–774.

Scheufele, D. A.; and Krause, N. M. 2019. Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, 116(16): 7662–7669.

Schlichtkrull, M.; Chen, Y.; Whitehouse, C.; Deng, Z.; Akhtar, M.; Aly, R.; Guo, Z.; Christodoulopoulos, C.; Cocarascu, O.; Mittal, A.; Thorne, J.; and Vlachos, A., eds. 2024. *Proceedings of the Seventh Fact Extraction and VERification Workshop (FEVER)*. Miami, Florida, USA: Association for Computational Linguistics.