

## Introduction to the Minitrack on Data Analytics, Data Mining, and Machine Learning for Social Media

Natalie Gerhart  
Creighton University  
[NatalieGerhart@creighton.edu](mailto:NatalieGerhart@creighton.edu)

Kevin D. Mentzer  
Nichols College  
[Kevin.Mentzer@nichols.edu](mailto:Kevin.Mentzer@nichols.edu)

David J. Yates  
Bentley University  
[DYates@bentley.edu](mailto:DYates@bentley.edu)

### Abstract

*Thank you for your interest in the Data Analytics, Data Mining, and Machine Learning for Social Media Minitrack. We continued to see strong interest in this minitrack with twenty-eight submissions of which twelve were accepted. This year's paper contributions vary in scope and focus and thus provide a wide-ranging view of the state of methodologies in examining the field of social media. In our introduction last year, we warned that "our digital communities continue to struggle with the challenge of being free and open" which has been highlighted with recent articles expressing concern that X/Twitter is no longer freely available for academic research. As such, we are heartened at the diversity of data sources that this year's papers represent. We thank all those who have submitted to this minitrack over the years along with everyone who continues to better understand social media's impact on our society. Once again, this year's papers provide many new insights into this vital area of research.*

### 1. Papers at a glance

As in prior years, our community shows global diversity with authors from Brazil, China, Finland, France, Germany, Italy, and the United States. The sources of data are also diverse this year with not only the usual X/Twitter papers, but also papers that used data from CNN, Daily Mail, TripAdvisor, Airbnb, Telegram, Weibo, Facebook, and Reddit.

We would like to begin with our nomination for best paper. Pawlik and Pan (2024) explored the rapidly expanding area of chatbots by studying conversations in Reddit communities in the paper *Inferences from Social Media Conversations about the Adoption of Chatbots*. The paper builds on our understanding of chatbot adoption by drawing on the Unified Theory of Acceptance and Use of Technology 2, applying deep learning textual analysis in this context. Social bots were also the focus of the work by Wu, Zhang, and Chen (2024), *What if Social Bots Be My Friends? Estimating*

*Causal Effect of Social Bots Using Counterfactual Graph Learning*. In this paper the authors investigate the influence level of social bots versus their human counterparts.

Fake news and how we detect false information continues to be an area of interest. The literature review titled *Fake News Detection by Machine Learning in Latin America: A Systematic Review* by Ngomo, Torres de Paiva, and Garcia (2024) recognizes that frequently research has a North American or European focus and thus surveys how researchers have examined this topic in Latin America. By better understanding regional nuances, researchers can better detect fake news. Rieskamp, Mirbabaie, Langer, and Kocur (2024) looked at the spread of fake news related to COVID-19 in *From Virality to Veracity: Examining False Information on Telegram vs. Twitter*. Their work highlights the impact opinion leaders have on both communities and demonstrates the ease by which information spreads particularly on the Telegram platform. Similar findings related to the spread of misinformation can be found in the paper by Gruzd, Ghenai, and Mai (2024) titled *How COVID-19 Conspiracy Theories Spread on Twitter*. This work found that automated accounts and accounts that have since been suspended played a significant role in the spread of misinformation. This study highlights the need of social media sites to be aggressive in the monitoring of accounts to reduce the harmful impact that results from the spread of misinformation.

Two papers explored dark sides of the human condition: depression and war. Kuang, Xie, and Yan (2024) tackle the topic of mental illness and the detection of depression using social media. Their Multi-Scale Temporal Prototype Network (MSTPNet) outperforms other state of the art methods by analyzing the content and temporal distribution of posts in the Weibo User Depression Detection Dataset. Specifically, this study assesses an interpretable depression detection model based on symptom-based depression diagnostic criteria. Amjadi and John (2024) used Moral Foundations Theory to better understand the discussion

around the conflict between Ukraine and the Russian Federation in *All is Fair in Love and War: Moral Foundations in English-Language Tweets during the First 36 Weeks of Conflict Between Ukraine and Russia*. This work shows that different regions used differing moral foundations to conceptualize and process the conflict. However, across all regions, tweets were dominated by content expressing loyalty, while content expressing fairness was infrequent.

Okpala, Rodriguez, Han, Meierhofer, Mammola, Halse, Kropczynski, and Johnson (2024) highlight that other species can be impacted by our social media use in their paper *A Framework for Perception Analysis of Social Media Data During Disease Outbreaks: Uncovering Patterns of Resentment Towards Bats*. This study illustrates the importance of understanding perceptions on social media by analyzing sentence structures and how meaning is drawn from words connected to their semantic networks.

In *Effectively Delivering Author's Point to Reader: Pointer-Generator Network Approach*, Kim and Hong (2024) propose and evaluate a summarization algorithm that combines elements of abstractive summarization with keyword extraction. Their work shows the potential effectiveness of summaries that encapsulate the authors' intentions in shaping readers' thoughts and behaviors.

Babu and Zhang (2024) use a deep learning framework to examine managers' responses and their relationship with product review convergence/divergence in *Systematic Contextual-based Affinity Analytics Research on Association of Manager Response and Customer Reviews*. This study examines the connections between manager and customer narratives over time and product price and reputation.

While most of the papers this year focused on text data, two papers used image data in their analysis. Crowe, Ricks, and Hall (2024) utilize a combination of text and image data in *Enhancing User Behavior Modeling via Machine Learning with Combined Text and Image Data*. Their work highlights the increased accuracy in predicting interaction metrics using combined data types. Finally, Feng, Li, and Zhang (2024), in *Visual Uniqueness: An Unsupervised Contrast Learning Approach*, take image use to a new level by showing that images that score high in uniqueness can lead to better product marketing.

## References

- Amjadi, E. & John, R. S. (2024). All is Fair in Love and War: Moral Foundations in English-Language Tweets during the First 36 Weeks of Conflict Between Ukraine and Russia. *Proceedings of the 57th Hawaii International Conference on System Sciences*.
- Babu, X. & Zhang, J. (2024). Systematic Contextual-based Affinity Analytics Research on Association of Manager Response and Customer Reviews. *Ibid*.
- Crowe, C., Ricks, B., & Hall, M. (2024). Enhancing User Behavior Modeling via Machine Learning with Combined Text and Image Data. *Ibid*.
- Feng, F., Li, C., & Zhang, S. (2024). Visual Uniqueness: An Unsupervised Contrast Learning Approach. *Ibid*.
- Gruzd, A., Ghenai, A., & Mai, P. (2024). How COVID-19 Conspiracy Theories Spread on Twitter. *Ibid*.
- Kim, H. & Hong, S. (2024). Effectively Delivering Author's Point to Reader: Pointer-Generator Network Approach. *Ibid*.
- Kuang, J., Xie, J., & Yan, Z. (2024). What Symptoms and How Long? An Interpretable AI Approach for Depression Detection in Social Media. *Ibid*.
- Ngomo, J. G. N., Torres de Paiva, R., & Garcia, A. C. B. (2024). Fake News Detection by Machine Learning in Latin America: A Systematic Review. *Ibid*.
- Okpala, I., Rodriguez, G. R., Han, C., Meierhofer, M., Mammola, S., Halse, S., Kropczynski, J., & Johnson, J. (2024). A Framework for Perception Analysis of Social Media Data During Disease Outbreaks: Uncovering Patterns of Resentment Towards Bats. *Ibid*.
- Pawlik, V. P. & Pan, Y. (2024). Inferences from Social Media Conversations about the Adoption of Chatbots. *Proceedings of the 57th Hawaii International Conference on System Sciences*.
- Rieskamp, J., Mirbabaie, M., Langer, M., & Kocur, A. (2024). From Virality to Veracity: Examining False Information on Telegram vs. Twitter. *Ibid*.
- Wu, Z., Zhang, Y., & Chen, X. (2024). What if Social Bots Be My Friends? Estimating Causal Effect of Social Bots Using Counterfactual Graph Learning. *Ibid*.