Toward a Capture-Track-Respond Framework: A Survey of Data-Driven Methods for Countering LLM-Generated Misinformation

Han Kyul Kim[1]*, **Andy Skumanich**[2]*
[1]University of Southern California
[2]Innov8ai
hankyulk@usc.edu, askuman@innov8ai.com

Abstract

The rapid rise of generative AI (GenAI) has revolutionized online communication while simultaneously fueling the proliferation of AI-generated misinformation. Despite safety protocols implemented by major GenAI providers, adversarial tactics and unmoderated platforms continue to facilitate the unchecked spread of harmful misinformation. Addressing this growing threat demands scalable, data-driven solutions. This paper introduces the Capture-Track-Respond (CTR) framework, which systematically integrates advanced AI techniques to identify, monitor, and counter misinformation. The Capture mode minimizes reliance on costly data annotation through approaches such as active learning and domain adaptation. The Track mode analyzes how misinformation evolves over time and across networks using time-series and network analysis, ensuring adaptability to dynamic environments. The Respond mode combines AI-driven insights with human expertise to develop precise and efficient countermeasures. By detailing the AI strategies underpinning each mode, this paper provides a comprehensive roadmap for deploying CTR to combat LLM-generated misinformation at scale. We aim to foster collaboration among researchers, technologists, and policymakers to safeguard the integrity of information ecosystems in the GenAI era.

Introduction

The rise of generative AI (GenAI), powered by large language models (LLMs), has rapidly transformed how people communicate and consume information. On one hand, LLMs deliver remarkable advances in natural language processing (NLP) tasks, enabling highly accurate translation (Brants et al. 2007; Zhang, Haddow, and Birch 2023), text summarization (Zhang et al. 2024), and content creation at previously unimaginable scales. On the other hand, these same breakthroughs open new avenues for misuse, particularly in the realm of misinformation (Skumanich and Kim 2024a; Chen and Shu 2024). When deployed maliciously, fine-tuned LLMs can produce vast quantities of deceptively human-like text designed to mislead readers. Recent studies indicate that such AI-generated misinformation can be even harder to detect than its human-crafted counterparts, for both everyday readers and state-of-the-art detection methods (Chen and Shu 2023). Furthermore, the ability to automate every step of the misinformation life cycle, from ideation to dissemination, amplifies these risks by enabling large-scale influence campaigns at an unprecedented speed.

In response to these growing challenges, this paper introduces a practical **Capture-Track-Respond (CTR)** framework that unifies previous research efforts into a comprehensive, deployable system capable of addressing misinformation in near real-time. While many existing techniques focus on isolated aspects of detection or response, the CTR framework we propose is designed to systematically integrate and automate them. Specifically, **Capture** targets the initial identification of misleading content, such as social media posts, while **Track** focuses on monitoring how misinformation evolves after its initial detection. Finally, **Respond** relies on human input, supported by AI-driven insights, to correct or counter the spread of false information.

This work draws heavily on lessons learned from the authors' experience developing misinformation capture systems during the COVID era (Skumanich and Kim 2024b,c; Kim and Skumanich 2024). By detailing the essential technical components for each stage of CTR, our goal is to illustrate how data and AI algorithms can be harnessed to create a scalable real-time misinformation defense system. In doing so, our goal is to bridge the gap between experimental research and large-scale, practical deployment, moving us closer to a safer information ecosystem in the age of GenAI.

Sources of LLM-generated Misinformation

A key motivation for a unified framework such as Capture-Track-Respond (CTR) lies in the sheer volume of online misinformation. This section briefly categorizes two primary sources of text-based misinformation produced by LLMs. Although our focus here is on text, the same principles can be extended to other data modalities, such as images and videos, given the emergence of powerful generative models for these formats (Li et al. 2018, 2019; Ho et al. 2022).

Misinformation from moderated LLM services: Moderated LLM services, such as ChatGPT or other popular commercial LLMs, typically include safeguard measures such as content filters, policy guidelines, and continuous monitoring to prevent harmful or misleading content. However, these safeguard measures are by no means infallible. Despite state-of-the-art safeguard measures, adversarial users

*These authors contributed equally.

can often bypass moderation mechanisms through careful prompt engineering. For example, Chen and Shu (2023) demonstrates that specific prompt instructions can facilitate Controllable Misinformation Generation (CMG) with a high success rate. Moreover, the generated misinformation often exhibits more deceptive styles than the human-generated misinformation, making detection challenging for both human readers and automated detectors. Since these platforms are widely available and easily accessible, even a partial circumvention of their safeguards can produce large volumes of misleading content. For a more detailed analysis of strategies used to generate or detect misinformation in moderated LLM services, readers may refer to Chen and Shu (2023).

Misinformation from unmoderated LLM service: In contrast, unmoderated or open-source LLMs are often deployed on fringe social networks that either reject or minimize content moderation (Skumanich and Kim 2024b). Motivated by concerns about "free speech" or distrust of mainstream platforms, these networks allow users to exploit LLMs to generate vast amounts of misinformation with few or no guardrails in place.

Furthermore, unmoderated LLMs can be fine-tuned on narrowly targeted or biased datasets sourced from these fringe communities, enabling them to produce content aligned with extreme ideologies and reinforcing echo chambers. This unrestricted access permits large-scale, highly targeted misinformation campaigns. In some cases, these communities can accelerate the spread of false narratives, as they are shared among like-minded members eager to embrace and circulate ideologically charged information. The potential for viral misinformation is especially high in these spaces, given the absence of any policy enforcement system. Consequently, these platforms can generate or host persuasive AI-driven content that remains unchallenged and is often consumed by the same user base, ultimately undermining the credibility of online information.

Capture-Track-Respond Framework

Given the sheer volume and coherence of misinformation that different LLM sources can generate, it is clear that robust, data-driven methods are essential to confront this increasing societal challenge. However, relying on any single algorithm or technique falls short in the real world, where the topics and formats of misinformation change dynamically across multiple online platforms. For example, a method designed to detect COVID-19 misinformation might fail to identify newly emerging political conspiracies or fabricated news stories that adopt different linguistic styles, formats, or cultural references.

To address the rapidly shifting nature of online misinformation at a scale, this paper introduces a unified Capture-Track-Respond (CTR) framework. This framework incorporates machine learning, data mining, and NLP methods whose seamless integration is designed to scale and adapt to evolving misinformation trends. Figure 1 provides a high-level overview of the proposed CTR framework. As we will outline in subsequent sections, each component in CTR (Capture, Track, and Respond) targets a specific stage in the misinformation lifecycle. By doing so, the framework not only detects misinformation as it appears but also monitors how it evolves and provides actionable insights to humans or automated systems tasked with countering it. This multifaceted approach is crucial for keeping pace with the increasingly sophisticated tactics deployed by malicious actors and the continuous development of new LLM-based models.

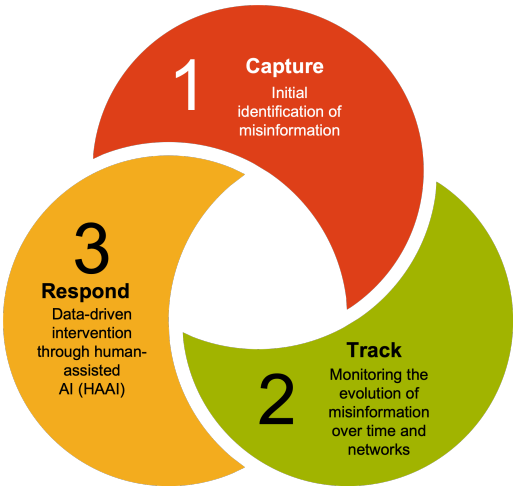

Figure 1: A graphical summary of the proposed CTR framework

Capture

The Capture mode of the CTR framework focuses on the initial detection of misinformation from dynamic data flows, such as social media posts, news articles, or other online content streams. This step is critical as it serves as the first line of defense against the spread of misinformation. While previous works (Shu et al. 2017; Islam et al. 2020; Abdali, Shaham, and Krishnamachari 2024) have proposed various machine learning algorithms for detecting misinformation, many fail to account for the rapidly evolving landscape of misinformation. As new topics, formats, and platforms emerge, systems that rely on static datasets and the assumptions of supervised learning often become outdated and less effective.

Although supervised approaches perform well in controlled environments, their reliance on human-annotated data presents significant scalability challenges. The cost and time required for annotation hinder their ability to keep up with the rapid generation of misinformation, particularly in diverse and constantly shifting domains. To address these limitations, the Capture mode builds upon existing misinformation detection methods by integrating data-driven techniques aimed at reducing dependence on annotated datasets,

Active learning

A key ingredient for reducing the reliance on annotated datasets is minimizing the amount of labeled data required for training. Active learning is a promising technique that facilitates this by optimizing the cost of detection model

training. It selectively queries the most informative unlabeled data points for human annotation, reducing the need for extensive labeled datasets. Rather than requiring a fully annotated dataset upfront, active learning algorithms iteratively identify and request labels for data samples expected to provide the most significant improvement in model performance (Figure 2). By prioritizing the most uncertain or representative samples, active learning minimizes the annotation burden while maximizing the model's effectiveness, making it an efficient alternative to traditional supervised learning methods.

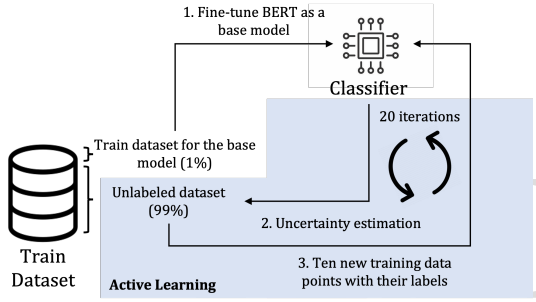

Figure 2: A summary of underlying mechanism of active learning from Kim and Skumanich (2024)

Its use in optimizing annotation processes for misinformation detection has shown significant potential. For example, Hasan, Alam, and Adnan (2020) was one of the first studies in misinformation detection to incorporate an active learning framework for training a political misinformation detector. Their findings demonstrated that high detection accuracy could be achieved with as little as 4% of the dataset annotated for fake news detection. Subsequent works have confirmed the effectiveness of active learning in this domain using various approaches, including convolutional neural networks (Bhattacharjee, Talukder, and Balantrapu 2017), large language models (Folino et al. 2024), and graph neural networks (Ren et al. 2020; Barnabò et al. 2023). However, these studies primarily frame misinformation detection through the lens of fake news, which represents only a small subset of the broader misinformation landscape.

As social media increasingly shapes how people consume and share information, only a limited number of studies have specifically applied active learning to misinformation detection in social media contexts. For example, Farinneya et al. (2021) investigated active learning for rumor detection on social media, while Kim and Skumanich (2024) proposed an innovative active learning method for detecting misinformation in short tweets, demonstrating its effectiveness.

Despite these advancements, existing research has exclusively focused on human-generated misinformation, leaving significant gaps in understanding whether these approaches can effectively address LLM-generated misinformation. The challenge is compounded by the diverse origins and types of misinformation that LLMs can generate, which introduce new complexities not addressed in prior studies. Consequently, developing and evaluating active learning techniques tailored to LLM-generated misinformation is a crucial first step in minimizing reliance on human annotation, a critical requirement for the success of the Capture mode in the CTR framework.

Domain adaptation

While active learning techniques are essential for reducing the annotation burden, their effectiveness is largely limited to in-domain misinformation detection. When the source or content of misinformation changes significantly, a new annotated dataset is often required to train additional detection models. Although active learning can be re-applied in such cases, this approach introduces delays, as the process of selecting and annotating data in the new domain takes time. These delays can hinder the timely deployment of detection models for emerging types of misinformation, reducing the ability to achieve early detection when it is most critical.

It is important to distinguish domain adaptation from transfer learning, which is often explored in the context of multi-modal (Singhal et al. 2020; Liu et al. 2023) or multi-lingual (Lee and Kim 2022; Ghayoomi and Mousavian 2022; Ozcelik et al. 2023; Hussain et al. 2024) misinformation detection. While transfer learning reuses an existing model by fine-tuning it on a small sample of annotated data from the new domain, domain adaptation stands apart by requiring no additional annotations for the target domain. This key difference makes domain adaptation particularly valuable for real-world deployment, as it allows the Capture mode to be implemented immediately when new misinformation emerges.

In the detection of human-generated misinformation, techniques such as adversarial training (Ng et al. 2023), contrastive learning (Yue et al. 2022; Zeng et al. 2024), and disentangled representation learning (Zhang et al. 2020) have been widely explored and proven effective. However, these methods were developed and validated primarily for human-generated misinformation, leaving their applicability to LLM-generated misinformation an area of active investigation.

Recent research has begun to address this gap. For instance, Beigi et al. (2024) investigated the use of contrastive learning for the model attribution task, specifically in the context of LLM-generated disinformation. While this represents a promising step forward, the broader application of domain adaptation techniques in mitigating LLM-generated misinformation will require further validation. This involves assessing these methods across diverse definitions of domains, including the specific LLMs used, the sources of LLM-generated misinformation, the content and context of the misinformation, and the varying types of misinformation being propagated.

To effectively address the challenges posed by LLM-generated misinformation, future research must expand the scope of these technical methods, ensuring their robustness and adaptability in handling the unique characteristics of this rapidly evolving threat.

Track

The effectiveness of the Track module relies heavily on the volume and diversity of the live data streams from which

misinformation is believed to originate or propagate. These sources can range from mainstream social media platforms to fringe social networks. A key challenge in scaling this module is securing robust and reliable data interfaces to these sources. Since the availability of APIs or web crawling capabilities varies significantly between platforms, significant engineering efforts are required to establish and maintain access. Additionally, effective data storage and management systems are essential to ensure that misinformation elements identified by the Capture module can be continuously monitored and analyzed within the Track module.

Tracking misinformation also involves understanding its evolution across two critical dimensions: temporal and network aspects. **Temporal evolution** examines how misinformation changes over time, revealing patterns such as shifts in narratives, peaks in activity, and the duration of influence within communities. **Network evolution**, on the other hand, analyzes how misinformation spreads within and across specific subsets of the population, focusing on propagation speed, dissemination patterns, and the influence of key users or groups. Together, these perspectives provide a comprehensive understanding of misinformation dynamics, enabling the Track module to monitor its progression and inform targeted interventions.

For instance, Green et al. (2021) conducted hourly trend time-series analysis on the varying volumes of different types of misinformation on Twitter, uncovering key patterns in how misinformation peaks and wanes over short periods. Similarly, Skumanich and Kim (2024c) and Skumanich and Kim (2024b) explored changes in hashtag distributions and keyword streamographs over time, offering insights into how specific narratives evolve or gain traction (Figure 3). These studies highlight not only the importance of temporal analysis but also its potential to uncover previously unnoticed trends, such as the coordinated timing of misinformation campaigns or the cyclical nature of certain narratives. Temporal analysis, when combined with advanced visualization techniques, can transform vast, noisy datasets into actionable insights, allowing researchers to identify trends in real time and predict the trajectory of emerging misinformation narratives.

In the context of network analysis, Shu, Bernard, and Liu (2019) provides a detailed classification of different network structures commonly used to detect the dissemination of misinformation, offering a foundation for understanding how information flows within digital spaces. Building on these technical insights, empirical studies have demonstrated the value of network analysis in uncovering the deeper dynamics of misinformation spread. For example, Shao et al. (2018) analyzed diffusion networks comprising two million tweets from the 2016 US Presidential Election, revealing strong segregation of information. Similarly, Luo, Cai, and Cui (2021) employed network visualization to analyze misinformation dissemination on Weibo, while Duzen, Riveni, and Aktas (2024) visualized community networks to examine the dynamics of COVID-19 misinformation on Twitter. These examples showcase the potential of network analysis to reveal the structural patterns that drive the spread of false information.

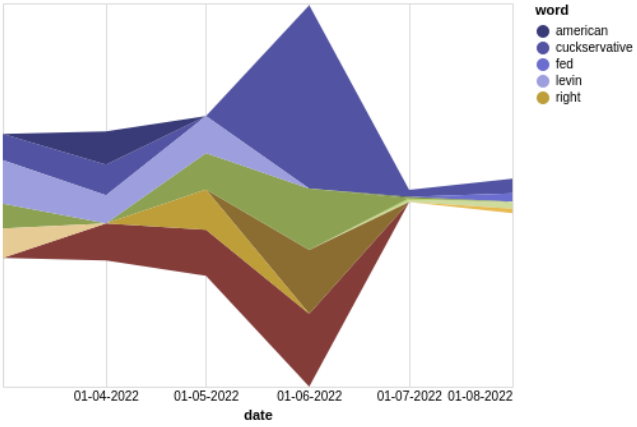

Figure 3: An examples of time-series visualization (Streamgraph) from Skumanich and Kim (2024b)

Building on these foundational works, LLM-generated misinformation introduces new complexities but also offers the potential for similarly actionable insights. Stable, long-term data sources capturing the diverse origins of LLM-generated misinformation, ranging from different LLM models, content types, and dissemination platforms, are critical to realizing this potential. By applying temporal and network perspectives, researchers can better understand the evolution of LLM-generated misinformation, quantify concerns over its rapid propagation, and identify critical intervention points. These insights will not only help confirm the scale of the threat posed by LLM-generated misinformation but also enable the development of more effective, data-driven strategies to mitigate its impact and preserve the integrity of information ecosystems.

Respond

Once practical and effective data-driven methods are implemented in the Capture and Track modes, the Respond mode focuses on mitigating the spread of misinformation and countering already disseminated content. While actual response activities will inevitably involve human intervention, integrating data-driven methods can significantly enhance the efficiency and precision of these efforts. The core of this integration lies in the concept of **human-assisted AI (HAAI)**, which combines human intuition and expertise with the scalability and analytical power of AI. This synergy allows AI techniques to augment human decision-making by providing actionable insights and streamlining the response process.

The adoption of HAAI in misinformation response is particularly relevant for tasks such as crafting targeted counter-narratives, prioritizing interventions, and identifying key influencers within misinformation networks. Techniques such as keyword extraction (Bae et al. 2021) and keyness analysis (Hardie 2014) enable precise identification of the themes, phrases, and language styles associated with misinformation, helping human operators tailor their responses to the most critical aspects of the issue. For example, keyword ex-

traction can highlight emerging narratives or repeated patterns across misinformation content, while keyness analysis can identify distinguishing linguistic features between misinformation and legitimate information sources (Skumanich and Kim 2024c).

Beyond these technical methods, HAAI also supports the development of broader strategies by quantifying the effectiveness of different response techniques. This includes evaluating how misinformation dynamics evolve after a countermeasure is deployed, enabling a feedback loop that refines future responses. Although trust in AI-driven results is an ongoing area of research, these techniques can provide valuable support by reducing the cognitive load on human operators, ensuring responses are data-informed, and enhancing the overall speed and effectiveness of mitigation efforts.

The Respond mode, underpinned by HAAI, serves as a critical bridge between automated systems and human oversight. By leveraging AI-driven insights and optimizing human-AI collaboration, this mode not only mitigates the immediate effects of misinformation but also builds a foundation for more resilient and adaptive response frameworks capable of addressing the evolving landscape of LLM-generated misinformation.

Implication of CTR for Societal Well-being

Misinformation poses a significant threat to societal well-being by spreading false narratives that erode public trust, fuel fear, and deepen social divisions. This impact is especially pronounced during crises, such as public health emergencies or political events, where misinformation can exacerbate panic or mislead critical decision-making. The CTR framework offers a scalable, data-driven solution to this challenge by systematically addressing the lifecycle of misinformation. By reducing the prevalence of misinformation and empowering stakeholders with actionable insights, the CTR framework fosters a healthier, more informed, and resilient society that can better navigate the complexities of the digital age.

Furthermore, it embodies the principles of human-compatible AI by integrating HAAI within its Respond mode. Recognizing the irreplaceable value of human expertise and judgment, our framework positions AI as a tool to augment, rather than supplant, human decision-making. By maintaining human oversight, our framework ensures that the decision-making process remains aligned with societal values, ethical considerations, and cultural sensitivities.

Conclusion

The proliferation of LLM-generated misinformation poses an unprecedented threat to the integrity of online communication and societal trust. As generative AI continues to scale in capability and accessibility, the risks associated with its misuse grow exponentially, demanding immediate and innovative solutions. This paper introduces the CTR framework as a systematic and scalable approach to address this critical challenge. By integrating advanced AI techniques, CTR offers a comprehensive methodology to identify, monitor, and counter the dynamic nature of misinformation.

The urgency of this challenge cannot be overstated. Without immediate and scalable interventions, the continued evolution of LLM-generated misinformation risks eroding public trust, deepening social divides, and amplifying harm on a global scale. Through this work, we aim to inspire the development and deployment of proactive, data-driven systems that not only combat misinformation but also reinforce the resilience of digital platforms in the GenAI era.

References

Abdali, S.; Shaham, S.; and Krishnamachari, B. 2024. Multi-modal misinformation detection: Approaches, challenges and opportunities. *ACM Computing Surveys*, 57(3): 1–29.

Bae, Y. S.; Kim, K. H.; Kim, H. K.; Choi, S. W.; Ko, T.; Seo, H. H.; Lee, H.-Y.; and Jeon, H. 2021. Keyword extraction algorithm for classifying smoking status from unstructured bilingual electronic health records based on natural language processing. *Applied Sciences*, 11(19): 8812.

Barnabò, G.; Siciliano, F.; Castillo, C.; Leonardi, S.; Nakov, P.; Da San Martino, G.; and Silvestri, F. 2023. Deep active learning for misinformation detection using geometric deep learning. *Online Social Networks and Media*, 33: 100244.

Beigi, A.; Tan, Z.; Mudiam, N.; Chen, C.; Shu, K.; and Liu, H. 2024. Model attribution in llm-generated disinformation: A domain generalization approach with supervised contrastive learning. In *2024 IEEE 11th International Conference on Data Science and Advanced Analytics (DSAA)*, 1–10. IEEE.

Bhattacharjee, S. D.; Talukder, A.; and Balantrapu, B. V. 2017. Active learning based news veracity detection with feature weighting and deep-shallow fusion. In *2017 IEEE International Conference on Big Data (Big Data)*, 556–565. IEEE.

Brants, T.; Popat, A.; Xu, P.; Och, F. J.; and Dean, J. 2007. Large language models in machine translation. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, 858–867.

Chen, C.; and Shu, K. 2023. Can llm-generated misinformation be detected? *arXiv preprint arXiv:2309.13788*.

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

Duzen, Z.; Riveni, M.; and Aktas, M. S. 2024. Analysing Impact Dynamics of Misinformation Spread on X (formerly Twitter) with a COVID-19 Dataset. *IEEE Access*.

Farinneya, P.; Pour, M. M. A.; Hamidian, S.; and Diab, M. 2021. Active learning for rumor identification on social media. In *Findings of the association for computational linguistics: EMNLP 2021*, 4556–4565.

Folino, F.; Folino, G.; Guarascio, M.; Pontieri, L.; and Zicari, P. 2024. Towards Data-and Compute-Efficient Fake-News Detection: An Approach Combining Active Learning and Pre-Trained Language Models. *SN Computer Science*, 5(5): 470.

Ghayoomi, M.; and Mousavian, M. 2022. Deep transfer learning for COVID-19 fake news detection in Persian. *Expert Systems*, 39(8): e13008.

Green, M.; Musi, E.; Rowe, F.; Charles, D.; Pollock, F. D.; Kypridemos, C.; Morse, A.; Rossini, P.; Tulloch, J.; Davies, A.; et al. 2021. Identifying how COVID-19-related misinformation reacts to the announcement of the UK national lockdown: An interrupted time-series study. *Big Data & Society*, 8(1): 20539517211013869.

Hardie, A. 2014. Log ratio: An informal introduction. *ESRC Centre for Corpus Approaches to Social Science (CASS)*, 1–2.

Hasan, M. S.; Alam, R.; and Adnan, M. A. 2020. Truth or lie: Pre-emptive detection of fake news in different languages through entropy-based active learning and multi-model neural ensemble. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 55–59. IEEE.

Ho, J.; Chan, W.; Saharia, C.; Whang, J.; Gao, R.; Gritsenko, A.; Kingma, D. P.; Poole, B.; Norouzi, M.; Fleet, D. J.; et al. 2022. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*.

Hussain, A.; Nawabi, A. K.; Alam, M.; Iqbal, M. S.; and Hussain, S. 2024. Detecting Urdu COVID-19 misinformation using transfer learning. *Social Network Analysis and Mining*, 14(1): 143.

Islam, M. R.; Liu, S.; Wang, X.; and Xu, G. 2020. Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining*, 10(1): 82.

Kim, H. K.; and Skumanich, A. 2024. Active Learning for Practical Misinformation Classification in Social Media: a Case Study on COVID-19. In *2024 IEEE International Conference on Big Data (BigData)*. IEEE.

Lee, J.-W.; and Kim, J.-H. 2022. Fake sentence detection based on transfer learning: applying to Korean COVID-19 fake news. *Applied Sciences*, 12(13): 6402.

Li, B.; Qi, X.; Lukasiewicz, T.; and Torr, P. 2019. Controllable text-to-image generation. *Advances in neural information processing systems*, 32.

Li, Y.; Min, M.; Shen, D.; Carlson, D.; and Carin, L. 2018. Video generation from text. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Liu, H.; Wang, W.; Sun, H.; Rocha, A.; and Li, H. 2023. Robust domain misinformation detection via multi-modal feature alignment. *IEEE Transactions on Information Forensics and Security*.

Luo, H.; Cai, M.; and Cui, Y. 2021. Spread of misinformation in social networks: Analysis based on Weibo tweets. *Security and Communication Networks*, 2021(1): 7999760.

Ng, K. C.; Ke, P. F.; So, M. K.; and Tam, K. Y. 2023. Augmenting fake content detection in online platforms: A domain adaptive transfer learning via adversarial training approach. *Production and Operations Management*, 32(7): 2101–2122.

Ozcelik, O.; Yenicesu, A. S.; Yildirim, O.; Haliloglu, D. S.; Eroglu, E. E.; and Can, F. 2023. Cross-Lingual Transfer Learning for Misinformation Detection: Investigating Performance Across Multiple Languages. In *Proceedings of the 4th Conference on Language, Data and Knowledge*, 549–558.

Ren, Y.; Wang, B.; Zhang, J.; and Chang, Y. 2020. Adversarial active learning based heterogeneous graph neural network for fake news detection. In *2020 IEEE International Conference on Data Mining (ICDM)*, 452–461. IEEE.

Shao, C.; Hui, P.-M.; Wang, L.; Jiang, X.; Flammini, A.; Menczer, F.; and Ciampaglia, G. L. 2018. Anatomy of an online misinformation network. *Plos one*, 13(4): e0196087.

Shu, K.; Bernard, H. R.; and Liu, H. 2019. Studying fake news via network analysis: detection and mitigation. *Emerging research challenges and opportunities in computational social network analysis and mining*, 43–65.

Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1): 22–36.

Singhal, S.; Kabra, A.; Sharma, M.; Shah, R. R.; Chakraborty, T.; and Kumaraguru, P. 2020. Spotfake+: A multimodal framework for fake news detection via transfer learning (student abstract). In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 13915–13916.

Skumanich, A.; and Kim, H. K. 2024a. Modes of Analyzing Disinformation Narratives With AI/ML to Assist in Mitigating the Weaponization of Social Media. In *Workshop Proceedings of the 18th International AAAI Conference on Web and Social Media*.

Skumanich, A.; and Kim, H. K. 2024b. Modes of tracking mal-info in social media with AI/ML tools to help mitigate harmful GenAI for improved societal well being. In *Proceedings of the AAAI Symposium Series*, volume 3, 412–417.

Skumanich, A.; and Kim, H. K. 2024c. Time Series Analysis of Key Societal Events as Reflected in Complex Social Media Data Streams. *arXiv preprint arXiv:2403.07090*.

Yue, Z.; Zeng, H.; Kou, Z.; Shang, L.; and Wang, D. 2022. Contrastive domain adaptation for early misinformation detection: A case study on covid-19. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2423–2433.

Zeng, H.; Yue, Z.; Shang, L.; Zhang, Y.; and Wang, D. 2024. Unsupervised domain adaptation via contrastive adversarial domain mixup: A case study on covid-19. *IEEE Transactions on Emerging Topics in Computing*.

Zhang, B.; Haddow, B.; and Birch, A. 2023. Prompting large language model for machine translation: A case study. In *International Conference on Machine Learning*, 41092–41110. PMLR.

Zhang, H.; Qian, S.; Fang, Q.; and Xu, C. 2020. Multimodal disentangled domain adaption for social media event rumor detection. *IEEE Transactions on Multimedia*, 23: 4441–4454.

Zhang, T.; Ladhak, F.; Durmus, E.; Liang, P.; McKeown, K.; and Hashimoto, T. B. 2024. Benchmarking large language

models for news summarization. *Transactions of the Association for Computational Linguistics*, 12: 39–57.