# Research Statement

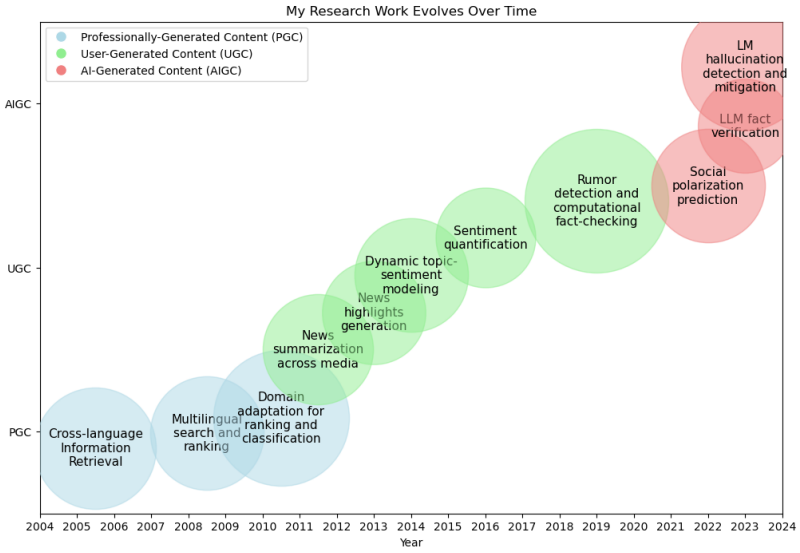
Wei Gao
School of Computing and Information Systems, Singapore Management University
Tel: (65) 6826-1369; Email: weigao@smu.edu.sg
15 December 2025 (Updated)

## 1 Background and Strategy

Natural language content is a manifestation of human ingenuity and creativity, influencing and propelling the advancement of human civilization. If Artificial Intelligence (AI) stands as a crowning achievement, Natural Language Processing (NLP) represents the most illustrious jewel in this crown. My professional passion lies in leveraging computational methods to aid users and applications in comprehending, recognizing, and generating valuable insights from natural language data and phenomena.

Since the turn of the century, there have been two transformative shifts in text content generation. The first emerged in early 2000s driven by the advent of social media while traditional professional content creation still held sway. From the late 2000s, the paradigm shift became prevalent and obvious, ushering in the coexistence of user-generated content (UGC) alongside professionally generated content (PGC). This shift liberates ordinary individuals to actively participate in content creation. The second significant transformation has taken place more recently, driven by rapid advancements in generative AI algorithms, particularly in the field of language modeling technologies. As a result, an increasing proportion of content is now being created with machine assistance or fully automated generation, namely AI generated content (AIGC). This trend points towards a future where Internet content and applications will continue to diversify and become increasingly virtualized.

During the past decade, my research has been conducted within this evolving landscape. My primary focus has been on designing and developing natural language technologies and applications to address the challenges of information processing through innovative and forward-looking thinking, data-driven approaches based on machine learning, and empirical experimentation methodologies. A key aspect of my strategy is to consider both current and future developments, always guided by practical applications. The figure at the right hand side illustrates the progression of my research themes over time, closely aligned with the shifts in content creation and consumption demands.




## 2 My Research Development

### 2.1 Era of Professionally Generated Content (PGC)

During the early stage of my research career (2004-2009), the media landscape was largely homogeneous, with traditional web pages serving as the predominant content format. Texts were primarily authored by professional producers, and social media platforms were in their nascent stages. During this phase, my research focused on processing texts from traditional media and developing more effective text retrieval methods, specifically for searching web contents. My doctoral research centered on web search and ranking for multilingual environment. It addressed the challenge of language barriers hindering monolingual users from accessing information in different languages

through search engines. I also expanded the research to address the problem of cross-domain vertical search and result ranking in an environment with imbalanced information resources, aiming to effectively retrieve useful information from low-resource domains.

To facilitate cross-language retrieval, traditional methods relied on query translation. However, automatic translation technology during that time was much less effective than today, leading to translated queries that did not align with the target language's conventions. To overcome this limitation, I developed an innovative cross-lingual query suggestion approach [1, 4]. This approach involved identifying candidate queries in the target language from extensive search query logs given the source-language query, for which I utilized linear regression to learn how to match queries cross-lingually to suggest the optimal target-language queries.

Furthermore, in a multilingual search environment, search results in different languages often came from distinct models, posing the challenge of how to fairly rank multilingual search results. To address this, I devised a novel joint ranking model based on restricted Boltzmann machines, a stochastic recurrent neural network [2]. This model leveraged document relevance in one language to help relevance estimation in different languages, thereby improving the probabilistic inference of the relevance of entire multilingual ranking list. This novel idea was further extended by utilizing bilingually relevant information to improve monolingual search ranking algorithms [3].

Inspired by these ground works, my focus shifted towards domain adaptation problems in web search during a transitional period after my Ph.D. study. I explored the adaptation of ranking models [5, 6, 7, 12] and classification models [10, 11] to low-resource or different vertical search domains. Although these studies used traditional PGC as data, they laid the foundation for my subsequent research that encompassed heterogeneous media content forms.

## 2.2 From PGC to Social Media

With the popularity of social media, the influence of User-Generated Content (UGC) has become increasingly evident. During the 2011 social revolution in the Middle East, I observed the intricate interplay between social media and traditional media and the significant roles they played in shaping the movement. This prompted me to concentrate my research on cross-media information connections and complementarity to provide more accurate and comprehensive insights on concerned social events and topics.

Between 2012 and 2014, I designed and developed a range of technologies, including joint topic extraction, aspects alignment, and complementarity measurements for cross-media event summarization [8]. Notably, I focused on harnessing the real-time nature of social media content and devised a method for generating news highlights from relevant posts on Twitter [13, 14]. The models used mainstream news highlights as training targets, learning how authors write news highlights based on social media posts, particularly for the cases where news content is underreported by mainstream media.

Moreover, in the context of societal development and transformation, a precise understanding of public opinion is crucial for governments, individuals, and communities. Existing opinion analysis technology, however, could not adequately and efficiently monitor the dynamic evolution of topics and sentiments. To address this, I worked with my collaborators to pioneer two innovative methods: a dynamic joint model for predicting and tracking changes in topics and sentiments [9], and a tweet sentiment quantification model for estimating the prevalence of opinions on the given topics [18, 19, 20].

As my research progressed, I found that unreliable or even false information is often output by the system, as a result of garbage-in-garbage-out dilemma of data-driven approaches, which is fatal to the credibility of downstream applications. I anticipated that the factuality of UGC might emerge as vital factors affecting the wide acceptance of natural language applications. Meanwhile, I cast my doubt on its trustworthiness due to the lack of systematic moderation in social media, necessitating a shift in my research focus towards information credibility. Since 2015, I have guided a student in developing methods for rumor detection. We introduced a time series model based on social context information [21], and then developed the first method based on recurrent neural networks for rumor detection [22]. That was around 2016 when the world was on the verge of so-called post-truth era coming along marked by global events like Brexit and President Trump's election, which solidified my belief on this research direction to go. Subsequently, we came up with the first tree kernel learning method that utilized information propagation structure [24] in 2017, followed by delving into various deep neural network models to address various issues on rumor detection. We published over ten high-quality papers on this topic, which featured five significant

innovations: 1) Introduced the first tree structured rumor detection model based on recursive neural networks [25, 29], followed with a tree transformer [30]. 2) Developed the first multi-task model capable of simultaneously detecting rumors and post-level stances [26], followed by the weakly supervised joint detectors of misinformation based on multiple instance learning [33, 35]. 3) Employed generative adversarial learning to enhance the robustness of rumor detection against the noise from information campaign [27, 34]. 4) Pioneered a model that utilized user trust networks to identify rumor spreaders [23]. 5) Enhanced early rumor detection by leveraging neural Hawkes process [31] and imitation agent learning [32] to automatically determine the earliest time when a stable decision can be confidently output, which are by far the only two approaches that consider prediction accuracy, earliness and stability, outperforming existing methods in all of the three aspects.

### 2.2.1 From UGC to AIGC

While social media has long been a breeding ground for rumors and misinformation, causing harmful effects on a significant scale, recent developments, especially the rapid rise of transformer-based Large Language Models (LLM) like ChatGPT since the end of 2022, have posed a new and potentially dangerous future of scaled-up manipulation with non-factual and divisive information.

In response to social polarization, I have sought interdisciplinary collaborations with cognitive psychologists to design and develop tools for analyzing and monitoring socially polarizing speech on social media platforms. Our goal is to implement natural language understanding tools to quantitatively represent deep user moral values for analysing and predicting user stances and online behaviours. Our preliminary findings suggest that features of human moral foundations can help predict not only stances of individual posts but also user positions across different events [36, 37]. Additionally, I am guiding students to model users' credulity or vulnerability to information, enabling early prediction of viral rumors and their potential reachability based on graphical neural multi-task learning methods [40]. This research direction aims to provide accurate insights to policy makers for more targeted analysis and intervention facing misinformation.

To address the factuality of AIGC, I have begun to evaluate the reliability of LLMs and explore how to leverage their strengths while mitigating their potential harmful effects in the context of fact verification. My goal is to develop medium-sized language models that are more controllable, adjustable, and interpretable than LLMs which suffer from black-box nature inherently limiting their applicability. We have made progress in language model-based fact verification: 1) Fine-tune medium-sized language models with minimal data by utilizing the logical consistency of facts as constraints [38]. 2) Propose a hierarchical step-by-step prompting method to guide LLMs finding required external evidence based on retrieval-augmented generation framework for more reliable fact verification [39]. 3) Leverage retrieved evidence to generate fact-check justifications with a novel few-shot retrieval-augmented language model, marking a significant advancement over in-context learning models like Flan-T5, Llama2, and retrieval-augmented models like Atlas [41]. 4) Strengthen evidence selection with reinforcement retrieval for boosting fact-checking effectiveness of blackbox LLMs [42]. 5) Enable a small MLLM (13B) to outperform GPT-4V on multimodal fact-checking by using synthetic data and propose model-agnostic data selection methods to bridge the distribution gap between synthetic and real-world data [43].

Beyond these, I also work with industry collaborators to enhance LLMs reasoning and efficiency. For example, we enhance chain-of-thought reasoning by fine-tuning LLMs with tree-of-thought search paths via Chain of Preference Optimization (CPO) [44] for improved problem-solving performance, while, more recently, reducing KV cache redundancies in long-context LLMs based on lightweight transformation of models such as LLaMA into hybrid variants through lazy layers—those focusing on recent or initial tokens, achieving efficient inference with minimal performance loss [45].

Although these research endeavors are still in their early stages, given their cutting-edge nature, they are expected to have a significant impact on misinformation research that ultimately benefits society as a whole by addressing the challenges posed by AIGC and promoting reliable use of language models.

# 3 Research Achievements, Findings and Impacts

My foremost objective is to develop natural language processing methods that align with present and future content characteristics and challenges, ultimately benefiting society through downstream applications. In this pursuit, I strive to strike a balance between addressing current practical needs and ensuring the generalizability and adaptability to future developments. As of December 2025, my research contributions have resulted in over 100 peer-reviewed publications, with a total citation rate over 8,300 and an H-index of 37 according to Google Scholar. These metrics attest to the significance and impact of my work in the field. My most important research outcomes and findings can be summarized in the following three key areas.

## 3.1 Misinformation Detection and Fact-Checking

The detrimental impact of misinformation, particularly evident during the COVID-19 pandemic, has caused significant harm to individuals and societies alike. Having been engaged in research in this area for some time, my work has contributed to the formulation of a series of widely discussed methods concerning the spread and detection of rumors and misinformation [21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 33, 35, 34, 38, 39, 40]. This series of research has made a substantial impact in the field, amassing over 5,000 citations. The datasets we constructed have been widely adopted, and the developed approaches have been comprehensively compared and evaluated by numerous subsequent works.

Central to our insights is the coexistence and intertwining of the propagation of rumors and people's pursuit of truth. The interplay of discussion, debate, and influence eventually leads to clarifying the veracity of information. Consequently, effectively capturing signals from various aspects of text, stance expression, and communication structure necessitates not only employing limited knowledge in the methods but also harnessing the collective wisdom of a great number of users. For this purpose, we endow the model with a comprehensive understanding of text, time, and structure, while injecting sufficient sensitivity to detect and analyze these signals effectively. Such an approach enables us to combat the dissemination of misinformation and uphold the integrity of information in an era dominated by the rapid exchange of ideas and communication via social media platforms.

We embarked on an exploration of various neural network models, aiming to design and experiment with effective attention mechanisms that could capture the distinctive multivariate features of rumors and misinformation. In our groundbreaking work on rumor detection based on recurrent neural networks [22], which was published in 2016, we focused on considering both timing and text content features while leveraging the abstract representation capabilities of multi-layer neurons. The resulting feature representation demonstrated significant improvements compared to traditional models. This pioneering research has gained significant amount of citations, attesting to its impact in the field.

A distinguishing feature of our approach is the comprehensive utilization of information propagation structure. To achieve this, we first employed a non-neural tree kernel model [24] to incorporate propagation features, and then a tree-structured recursive neural network [25], followed by incorporating a specially designed attention mechanism [29]. This mechanism enables the model to attend to the signals of veracity expressed by users via a bottom-up or top-down recursion, thereby enhancing the sensitivity and discrimination of the model. Our three papers promoting this approach have become a prevalent benchmark, and our ACM TIST paper [29] has been recognized as one of the five outstanding papers nominated by the Editor-in-Chief of the journal in 2022.

To address the lack of robustness due to widespread dissemination of rumors facilitated by robotic trolls, we were the first to use generative adversarial networks (GAN) for rumor detection by modeling the discussions surrounding rumors as an information campaign [27]. We then extended it by leveraging a transformer-based generator adversarial generation of posts, which significantly enhanced the robust detection on multiple rumor datasets [34].

## 3.2 Text Quantification for Tweet Sentiment Analysis

A major milestone in my public opinion analysis research is the development of quantitative estimation method for the sentiment prevalence of tweets [18, 19]. This algorithm directly estimates the prevalence of sentiment categories over the entire dataset by optimizing the class prevalence measure of each category, surpassing the traditional

classify-and-count approach for the first time. This achievement holds significant importance for public opinion analysis on social media, as conventional methods only focus on sentiment classification of individual instances, neglecting the estimation of overall popularity of certain opinions — a crucial goal for many applications.

The application of this method proves particularly relevant in scenarios such as elections or polls, where it is imperative to gauge the collective preference of the voters, rather than focusing on individual sentiments. This groundbreaking work was awarded as the best paper runner-up by IEEE/ACM ASONAM 2015 and secured a first rank in the SemEval tweet quantification task evaluation of that year.

### 3.3 Cross-Media News Event Linking and Multilingual Web Search

The dynamic interplay between social media and news media served as a source of inspiration, leading me to develop a joint topic model that utilizes information from these heterogeneous platforms to generate summaries of news events [8]. By leveraging this model, related information can be linked at a finer-grained level, capturing both the main subject and subordinate aspects of an event, resulting in mutually complementary summaries. Due to the practical value of this approach, an European patent has been granted.

In the domain of multilingual web search, my contributions lied in addressing cross-lingual challenges in search and ranking. I developed a joint ranking model based on restricted Boltzmann machines, which leverages document correlations in one language to estimate the relevance of documents in different languages [2]. This approach improves the probabilistic inference of the relevance ranking for the entire list of search results. In 2009, employing complex neural networks to tackle multilingual web search ranking problem was considered cutting-edge and pioneering. As a result of its innovation, this work was honored with the ECIR 2009 Best Student Paper Award.

## 4 Future Direction and Aspiration

The landscape of NLP and AI research is experiencing significant transformation with the rise of large language models. Many anticipate that the dawn of artificial general intelligence (AGI) is on the horizon. While the breakthrough capabilities of these models are evident, so are their inherent technical limitations. One area of concern lies in the biased influence of training data on the model's output and the inaccuracy and unreliability of its generation. Moreover, the misuse and abuse of related technologies pose serious threats to society, ranging from harmful medical fake news to election interference. Social science research indicates that human beings struggle to overcome inherent cognitive biases, while such generation technologies could be misused to manipulate cognition, incite conflicts, conduct ideological propaganda, amplify unfairness, disseminate misinformation, foster divisive speech, etc..

In response to these challenges, my research endeavors will expand to the field of language models for combating misinformation and facilitating fact-checking, with the primary objectives to enhance the capabilities of language models for mitigating their potential harm. This entails a few specific directions:

- Systematic evaluation of the capability and reliability of underlying models for factuality-related applications. While general task-oriented evaluations are common, fact verification demands stricter requirements for factual evidence and correct reasoning. Hence, I am to explore the fact-checking capabilities of the model, ranging from constructing evaluation data to designing evaluation measures and methods, considering the impact of training bias and hallucinations.

- Enabling the model to engage in "active thinking" by injecting natural language reasoning mechanisms. Currently, language models operate still based on text statistics, lacking capability of deliberate reasoning. To address this limitation, I intend to study how to equip the model with the ability for planning and reasoning in a structured manner. This includes aspects like searching and organizing relevant evidence, planning and deriving thinking trajectories, and generating reasonable explanations that align with the thinking process. Such enhancements are vital for reasoning-intensive tasks heavily relying on reliable evidence acquisition and inference. This direction will also help enhance the generalizability and trustworthiness of AI systems.

- To mitigate potential harm from unreliable AI systems, we need to enhance resilience in the external environment including targeted detection, prediction, and intervention. Identifying and assessing the hallucinatory and biased

content is the basis of offering further countermeasures. Furthermore, understanding user traits indicating who are susceptible to specific content is important. To achieve this, I have initiated interdisciplinary collaborations to study how to represent users' positions and psychological states using public textual and behavioral expressions. Quantifying emotional dependence on specific types of speech is essential in establishing data and developing models to analyze and predict divisive speech and groups, ultimately aiding in the defense against false and divisive content through personalized generation and recommendations.

# References


[1] W. Gao, C. Niu, J.Y. Nie, M. Zhou, J. Hu, K.F. Wong, and H.W. Hon. Cross-lingual Query Suggestion Using Query Logs of Different Languages. SIGIR 2007.

[2] W. Gao, C. Niu, M. Zhou, and K.F. Wong. Joint Ranking for Multilingual Web Search. ECIR 2009.

[3] W. Gao, J. Blitzer, M. Zhou, and K.F. Wong. Exploiting Bilingual Information to Improve (Monolingual) Web Search. ACL 2009.

[4] W. Gao, C. Niu, J.Y. Nie, M. Zhou, K.F. Wong, and Hsiao-Wuen Hon. Exploiting Query Logs for Cross-lingual Query Suggestion. ACM Transactions on Information Systems 28(2), Artical 6, 2010.

[5] W. Gao, P. Cai, K.F. Wong, and A. Zhou. Learning to Rank Only Using Training Data from Related Domain. SIGIR 2010.

[6] P. Cai, W. Gao, A. Zhou, and K.F. Wong. Query Weighting for Ranking Adaptation. ACL 2011.

[7] P. Cai, W. Gao, K.F. Wong, and Aoying Zhou. Relevant Knowledge Helps in Choosing Right Teacher: Active Query Selection for Ranking Adaptation. SIGIR 2011.

[8] W. Gao, P. Li, and K. Darwish. Joint Topic Modeling for Event Summarization across News and Social Media Streams. CIKM 2012.

[9] Y. He, C. Lin, W. Gao, and K.F. Wong. Dynamic Joint Sentiment-topic Model. ACM Transactions on Intelligent Systems and Technology, Volume 5, Issue 1, Article 6, 2013.

[10] P. Yang and W. Gao. Multi-view Discriminant Transfer Learning. IJCAI 2013.

[11] P. Yang and W. Gao. Information-theoretic Multi-view Domain Adaptation: A Theoretical and Empirical Study. Journal of Artificial Intelligence Research, 49:201-525, 2014.

[12] W. Gao and P. Yang. Democracy is Good for Ranking: Towards Multi-view Rank Learning and Adaptation in web search. WSDM 2014.

[13] Z. Wei and W. Gao. Utilizing Microblogs for Automatic News Highlights Extraction. COLING 2014.

[14] Z. Wei and W. Gao. Gibberish, Assistant, or Master? Using Tweets Linking to News for Extractive Single-document Summarization. SIGIR 2015.

[15] J. Li, W. Gao, Z. Wei, B. Peng, and K.F. Wong. Using Content-level Structures for Summarizing Microblog Repost Trees. EMNLP 2015.

[16] K. Song, S. Feng, W. Gao, D. Wang, G. Yu, and K.F. Wong. Personalized Sentiment Classification based on Latent Individuality of Microblog Users. IJCAI 2015.

[17] J. Li, M. Liao, W. Gao, Y. He, and K.F. Wong. Topic Extraction from Microblog Posts Using Conversation Structures. ACL 2016.

[18] W. Gao and F. Sebastiani. Tweet Sentiment: From Classification to Quantification. ASONAM 2015.

[19] W. Gao and F. Sebastiani. From Classification to Quantification in Tweet Sentiment Analysis. Social Network Analysis and Mining, Volume:6, Issue:1, Article 19, 2016.

[20] G. Da San Martino, W. Gao, and F. Sebastiani. Ordinal Text Quantification. SIGIR 2016.

[21] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.F. Wong. Detect Rumors Using Time Series of Social Context Information on Microblogging Websites. CIKM 2015.

[22] J. Ma, W. Gao, P. Mitra, S. Kwon, B.J. Jansen, K.F. Wong, and Meeyoung Cha. Detecting Rumors from Microblogs with Recurrent Neural Networks. IJCAI 2016.

[23] B. Rath, W. Gao, J. Ma, and J. Srivastava. From Retweet to Believability: Utilizing Trust to Identify Rumor Spreaders on Twitter. ASONAM 2017.

[24] J. Ma, W. Gao, and K.F. Wong. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. ACL 2017.

[25] J. Ma, W. Gao, and K.F. Wong. Rumor Detection on Twitter with Tree-structured Recursive Neural Networks. ACL 2018.

[26] J. Ma, W. Gao, and K.F. Wong. Detect Rumor and Stance Jointly by Neural Multi-task Learning. WWW 2018 Companion.

[27] J. Ma, W. Gao, and K.F. Wong. Detect Rumors on Twitter by Promoting Information Campaigns with Generative Adversarial Learning. WWW 2019.

[28] J. Ma, W. Gao, S. Joty, and K.F. Wong. Sentence-level Evidence Embedding for Claim Verification with Hierarchical Attention Networks. ACL 2019.

[29] J. Ma, W. Gao, S. Joty, and K.F. Wong. An attention-based Rumor Detection Model with Tree-structured Recursive Neural Networks. ACM Transactions on Intelligent Systems and Technology, Vol. 11, Issue 4, Article No.:42, 2020.

[30] J. Ma and W. Gao. Debunking Rumors on Twitter with Tree Transformer. COLING 2020.

[31] F. Zeng and W. Gao. Early Rumor Detection Using Neural Hawkes Process with a New Benchmark Dataset. NAACL 2022.

[32] F. Zeng, Q. Shao, L. Cheng, W. Gao, S.F. Cheng, J. Ma, and C. Niu. LLM-based Few-Shot Early Rumor Detection with Imitation Agent. KDD 2026.

[33] R. Yang, J. Ma, H. Lin, and W. Gao. A Weakly Supervised Propagation Model for Rumor Verification and Stance Detection with Multiple Instance Learning. SIGIR 2022.

[34] J. Ma, J. Li, W. Gao, Y. Yang, and K.F. Wong. Improving Rumor Detection by Promoting Information Campaigns with Transformer-based Generative Adversarial Learning. IEEE Transactions on Knowledge and Data Engineering. Vol. 35, No. 03, 2023.

[35] R. Yang, W. Gao, J. Ma, H. Lin, and Z. Yang. WSDMS: Debunk Fake News via Weakly Supervised Detection of Misinforming Sentences with Contextualized Social Wisdom. EMNLP 2023.

[36] H. Zhang, H. Kwak, W. Gao, and J. An. Wearing Masks Implies Refuting Trump?: Towards Target-specific User Stance Prediction across Events in COVID-19 and US Election 2020. ACM WebSci 2023.

[37] H. Zhang, P. Bhattacharya, W. Gao, L. Wong, B. Loh, J. Simons, and J. An. Enhancing Stance Classification with Quantified Moral Foundations. ASONAM 2024.

[38] F. Zeng and W. Gao. Prompt to Be Consistent is Better than Self-consistent? Few-shot and Zero-shot Fact Verification with Pre-trained Language Models. Findings of ACL 2023.

[39] X. Zhang and W. Gao. Towards LLM-based Fact Verification on News Claims with a Hierarchical Step-by-Step Prompting Method. IJCNLP-AACL 2023.

[40] X. Zhang and W. Gao. Predicting Viral Rumors and Vulnerable Users with Graph-based Neural Multi-task Learning for Infodemic Surveillance. Information Processing and Management, Vol. 61, Issue 1, January 2024.

[41] F. Zeng and W. Gao. JustiLM: JustiLM: Few-shot Justification Generation for Explainable Fact-Checking of Real-world Claims. Transactions of the Association for Computational Linguistics, 12:334–354, 2024.

[42] X. Zhang and W. Gao. Reinforcement Retrieval Leveraging Fine-grained Feedback for Fact Checking News Claims with Black-Box LLM. LREC-COLING 2024.

[43] F. Zeng, W. Li, W. Gao, and Y. Pang. Multimodal Misinformation Detection by Learning from Synthetic Data with Multimodal LLMs. Findings of EMNLP 2024.

[44] X. Zhang, C. Du, T. Pang, Q. Liu, W. Gao, and M. Lin. Chain of Preference Optimization: Improving Chain-of-Thought Reasoning in LLMs. NeurIPS 2024.

[45] X. Zhang, F. Zhang, C. Du, C. Du, T. Pang, W. Gao, and M. Lin. LightTransfer: Your Long-Context LLM is Secretly a Hybrid Model with Effortless Adaptation. Transactions on Machine Learning Research. ISSN 2835-8856, 2025.