# Semantic Alignment of Malicious Question Based on Contrastive Semantic Networks and Data Augmentation (Abstract Reprint)

**Xinyan Wang[1], Jinshuo Liu[1], Juan Deng[1], Meng Wang[1], Qian Deng[1], Youcheng Yan[1], Lina Wang[1], Yunsong Ma[2], Jeff Z. Pan[3]**

[1]School of Cyber Science and Engineering, Wuhan University
[2]School of Computer Science, University of Sydney
[3]The University of Edinburgh, Edinburgh

**Abstract Reprint.** This is an abstract reprint of the journal article by Wang, Liu, Deng, Wang, Deng, Yan, Wang, Ma, and Pan (2025).

## Abstract

The identification and filtration of malicious texts in social media environments represent a significant technical challenge aimed at protecting users from online violence and disinformation. This complexity stems from the diversity and innovativeness of social media texts, which include unique expressions and special sentence structures. Particularly, malicious texts in interrogative forms pose alignment challenges with traditional corpora due to existing methods failure to exploit the texts deep global semantic representations. This issue is compounded by the scant research on Chinese texts, leading to inefficiencies in recognition accuracy. To mitigate these challenges, we introduce an innovative framework based on a Global Contrastive Semantic Network (GCSN), designed to enhance malicious text recognition efficiency and accuracy by deeply learning global semantic knowledge. It comprises an encoder for global semantic information modelling and a graph-matching network for semantic similarity evaluation between question pairs, enabling the accurate identification and filtering of malicious texts with complex structures. Furthermore, we introduce a semantic consistency-based data augmentation method (COMBINE), using real-world data to generate balanced positive and negative samples, enriching the dataset and enhancing the models ability to distinguish semantic consistency through contrastive learning. Experimental validation on two Chinese datasets demonstrates our models exceptional performance, affirming its applicationa value in social media malicious text recognition. Our code is available at https://github.com/Wxy13131313131/GCSN-COMBINE

## References

Wang, X.; Liu, J.; Deng, J.; Wang, M.; Deng, Q.; Yan, Y.; Wang, L.; Ma, Y.; and Pan, J. Z. 2025. Semantic Alignment of Malicious Question Based on Contrastive Semantic Networks and Data Augmentation. *Journal of Artificial Intelligence Research*, 82: 1243–1266.