Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

| RESEARCH ARTICLE

A Study of Ethical Issues in Natural Language Processing with Artificial Intelligence

Yongfeng Ma
Faculty of Humanities and Social Sciences, Dalian University of Technology, Dalian 116081, China
Corresponding Author: Yongfeng Ma, **E-mail**: mayongfeng2022@mail.dlut.edu.cn

| ABSTRACT

Natural language processing has started to be widely used in various fields after the development lag of the artificial language processing stage, statistical language processing stage, and deep learning stage. The ethical issues of natural language processing can no longer be ignored, and the research on the ethical issues involved in natural language processing has received corresponding attention. However, the close relationship between artificial intelligence and natural language processing has not been considered in past studies of natural language processing, and a separation between technology and ethics has emerged. The purpose of this paper is to summarize the current research on ethical issues of natural language processing in four aspects: predictability, privacy, decision and responsibility, and bias, respectively, from the relationship between AI and natural language processing in order to achieve a better understanding and prevention of ethical issues in the development of natural language processing with AI.

| KEYWORDS

Natural language processing; artificial intelligence; ethics

| ARTICLE INFORMATION

ACCEPTED: 03 March 2023 **PUBLISHED:** 19 March 2023 **DOI:** 10.32996/jcsts.2023.5.1.7

1. Introduction

1.1 Research Background

The development of Natural Language Processing (NLP) can be divided into the artificial language processing stage, the statistical language processing stage, and the deep learning stage. The artificial language processing stage refers to the early NLP mainly relied on manual rules and grammar to analyze and generate natural language by writing rules and dictionaries by hand. This phase mainly lasted until the 1980s; the statistical language processing phase refers to the gradual shift of NLP to statistical model-based approaches with the development of computer technology and the growing corpus. This phase mainly focused on the 1990s to the early 21st century, including the fields of machine translation, speech recognition and information retrieval; the deep learning phase refers to the development of deep learning technologies that have driven a new revolution in NLP in recent years, including the application of technologies such as convolutional neural networks, recurrent neural networks and attention mechanisms. These technologies have enabled NLP to make unprecedented progress in the fields of machine translation, automatic summarization, and sentiment analysis and have also given rise to many new applications, such as intelligent customer service and intelligent writing.

1.2 Problem Statement and Objectives

And with the development of natural language processing, the ethical issues involved have become more and more important, but past research has not understood AI and natural language processing together much, which would lose a macro view of natural language processing as AI technology. The relationship between artificial intelligence and natural language processing is close. Natural language processing is a subfield of artificial intelligence whose goal is to enable computers to understand and process human language and to achieve human-computer interaction and natural language communication. Natural language processing mainly involves text processing, speech recognition, speech synthesis, machine translation, information retrieval, automatic summarization, sentiment analysis, and many other aspects, many of which are the focus of AI research. In turn, artificial intelligence provides powerful support for natural language processing, especially the development of deep learning techniques, which has brought significant progress to natural language processing. Various algorithms and models of deep learning techniques, such as convolutional neural networks, recurrent neural networks, and attention mechanisms, have become the core technologies for natural language processing, providing effective solutions for tasks such as text classification, sentiment analysis, and machine translation.

So nowadays, with the rapid development of artificial intelligence (AI) technology, natural language processing (NLP) techniques are widely used in various fields. However, NLP techniques face many ethical issues that need our attention, especially when combined with deep learning techniques. These issues include privacy, discrimination and bias, predictability, liability and decision-making issues. These issues may lead to bias and injustice in algorithms, resulting in unfair treatment of certain people or groups, as well as misdiagnosis in the medical field. To prevent these problems from continuing to worsen, many scholars have researched and developed appropriate measures.

1.3 Methodology

The research method adopted in this paper is desk research. Desk research is a research method that conducts academic research or social investigation by organizing, analyzing, and synthesizing secondary sources such as literature, data, and statistics. Taking the research process of this paper as an example, the research first needs to screen out the sources that are highly relevant and credible to the ethical issues of artificial intelligence natural language processing and record the information from each source for subsequent citation. Next, the collected information needs to be classified, organized and archived. This will help in the subsequent finding and citation of the information. After completing the data organization, the collected information needs to be analyzed. The final step is to start writing the research paper based on factors such as research questions, data classification and analysis results. In this paper, desk research is used because, compared to other forms of research methods, it focuses more on the collection, organization, and analysis of secondary data, which has the advantages of low time cost and accurate sample matching and is more suitable for the study of ethical issues related to natural language processing of artificial intelligence.

2. Questions and Analysis

2.1 Predictability Issues

Xiao and Wang (2019) explored how to quantify uncertainty in natural language processing tasks. The authors argue that uncertainties in NLP tasks are usually caused by two aspects: the first is data uncertainty, which includes factors such as the polysemy of the language itself, lexical and syntactic irregularities, and the quantity and quality of the training data. The second is model uncertainty, which includes factors such as overfitting, underfitting, and insufficient generalization ability of the model. To measure these uncertainties, the authors propose a new framework called "TUNA" (Tasks with Uncertainties and NAtural language), which enables models to accurately assess their prediction reliability by embedding uncertainties into NLP task reliability. The authors also propose a probability-based approach to measure the uncertainty of the model, which uses a Monte Carlo method to sample from the model's predictions and calculate the variability between different predictions, as well as the variability between predicted and true labels. Finally, the authors experimentally demonstrate the effectiveness of the TUNA framework, proving that the method can more accurately assess model uncertainty and achieve better performance than existing methods in several NLP tasks.

2.2 Privacy Issues

Liu et al. (2021) mentioned that privacy is one of the key issues to be addressed in natural language processing. Some existing methods often rely on collecting and using users' sensitive data in order to obtain better performance, which poses a potential risk to users' privacy. In this paper, we propose a privacy-preserving natural language processing method based on Locality-Sensitive Hashing (LSH). The basic idea of the method is to transform the raw text data into a vector and encode the vector into a binary code using LSH to protect the privacy of the user data. The method also uses a noise addition technique to add some random noise to the encoded data to protect the privacy of the user data when learning word embeddings. The amount of noise addition is adaptively adjusted according to the privacy requirements and the characteristics of the data. The experimental results in this paper show that the method can obtain similar performance as the original method while preserving user privacy. This suggests that the method is an effective privacy-preserving scheme that can be used for data privacy preservation in natural language processing tasks.

Tang (2022) mentions that privacy is a natural language processing, an important challenge. Since natural language processing tasks usually involve collecting and processing sensitive user information, appropriate privacy protection measures must be taken to protect users' privacy. The paper presents an overview of existing privacy-preserving techniques in natural language processing. These techniques can be divided into two categories: differential privacy techniques and encryption techniques. The differential privacy technique is a technique that protects data privacy by adding noise. The technique can be applied to many tasks in natural language processing, such as text classification, clustering, embedding learning, etc. In this paper, we introduce some recent differential privacy techniques and evaluate their effectiveness on different natural language processing tasks. Cryptography is a technique to protect data privacy through encryption and decryption. The technique is often applied to protect privacy during data transmission, such as encrypted communication and encrypted search. This paper introduces some of the latest cryptographic techniques and explores the prospects of their application in natural language processing tasks. This paper also discusses the advantages and disadvantages of differential privacy techniques and encryption techniques and explores how to choose the most appropriate privacy-preserving techniques for different natural language processing tasks.

2.3 Responsibility and Decision Making Issues

In Responsible natural language processing: A principlist framework for social benefits, Behera et al. (2023) propose a principlist framework for responsible NLP development and use, which includes six principles: benefit, harmlessness, fairness, respect for autonomy, respect for privacy, and respect for intellectual property. The article explains each principle in detail and discusses how they can be applied to NLP.

In “Artificial intelligence and responsibility gaps: what is the problem?”, Königs (2022) explores the issue of responsibility gaps arising from artificial intelligence technologies. The authors argue that in the development of AI technologies, there is often a gap between technological implementation and ethical standards, i.e., the rapidity and scale of technological development make the development and practice of ethical standards unable to keep up with the speed of technological development. The authors point out that the existence of such a responsibility gap may lead to the application of AI technologies with negative effects on humans and society and therefore needs to be carefully studied and addressed. The article also explores some existing solutions, such as introducing stricter regulations and norms, enhancing social assessment of technologies and encouraging companies and governments to focus on ethical standards alongside technology development. The authors argue that to address the responsibility gap; multiple measures are needed to promote synergy between technology and ethics for the sustainable development of AI technologies.

In “Improving humanitarian needs assessments through natural language processing”, Kreutzer explores how natural language processing (NLP) techniques can improve the accuracy and efficiency of humanitarian needs assessments. The authors argue that NLP can help humanitarian organizations quickly and accurately understand the information in large amounts of data and help them better understand the needs of recipients. The authors first review current research on the use of NLP in the humanitarian field and point out that some challenges remain in this area, such as the need for scenario-specific training corpora, language diversity and translation, interpretability, and fairness. The authors then present a framework for using NLP to improve humanitarian needs assessment. The framework consists of the following steps: data preparation, corpus construction, feature extraction, classification, and evaluation. The authors also present the use of NLP techniques to address three specific problems in the humanitarian domain: text classification, sentiment analysis, and information extraction. Finally, the authors summarize the contributions of the study and discuss directions for future research. This research is important for promoting the humanitarian field and responding to various crises.

The topic of this article by Jin and Mihalcea (2022), ‘Natural Language Processing for Policymaking’, is the application of natural language processing (NLP) to policymaking. The authors first introduce the basic concepts of policymaking and then discuss the applications of NLP in the field, especially in policy analysis and policy implementation. The article also introduces some key NLP techniques, such as text classification, named entity recognition, and sentiment analysis, and explains how they can be applied to policy formulation. The authors also mention some challenges and limitations, such as the quality and quantity of data, linguistic diversity, and the interpretability of the techniques. Finally, the authors discuss the potential implications of NLP in policy-making, such as information asymmetry, privacy and fairness issues, and ethical issues that policymakers need to consider. Overall, this paper provides a basic framework on how NLP can be applied to policymaking and offers some relevant technical and ethical issues for the reader's reference.

2.4 Bias Issues

Hovy and Prabhumoye (2021), in Five sources of bias in natural language processing, identified five major sources of bias in natural language processing (NLP). First, data bias is due to an imbalance in training data, mislabeling or missing markers, and the limited nature of the dataset. Second, model bias is due to the limitations of the model, such as the chosen algorithm or model architecture. Third, assessment bias is due to the choice of assessment metrics or the limited nature of the assessment dataset. Fourth, algorithm bias is due to the design or selection of a particular algorithm or process. Finally, social bias is due to the fact that NLP systems are designed and applied to specific social contexts and purposes and thus may reflect or enhance real-world inequalities or discrimination. The paper proposes several approaches to mitigate these biases, such as data augmentation, model adaptation, redefinition of assessment metrics, algorithm improvement, and social engagement.

Garrido-Muñoz et al. (2021) provide an overview of biases in deep natural language processing in A Survey on Bias in Deep NLP. The article first describes the different types of bias present in natural language processing, including dataset bias, algorithm bias, evaluation bias, and user interaction bias. Then, the article delves into the root causes and effects of these biases and summarizes different approaches used in current research to identify and mitigate them. These approaches include data augmentation, model fine-tuning, adversarial training, multi-task learning, and fairness constraints. Finally, the paper identifies directions and challenges for future research, such as how to better assess and understand biases and how to develop appropriate policies and regulations to ensure the fairness of natural language processing.

Stanczak and Augenstein (2021) discuss the issue of gender bias in natural language processing (NLP) in A Survey on Gender Bias in Natural Language Processing. The article analyzes the gender biases in existing NLP techniques, including corpus, pre-trained models, text classification, sentiment analysis, and machine translation. The authors also explore methods to address these gender biases, such as data augmentation, transfer learning, adversarial training, and fair learning. In addition, the article discusses how gender data collection and privacy protection should be conducted and suggests directions for future research to better address gender bias. In conclusion, the article aims to draw attention to gender bias in NLP and to provide a reference for further research.

3. Conclusion

The unpredictability of NLP models stems from several aspects. The text problem is due to the difficulty in predicting the behavior of the model due to the variety and constant change of text data that NLP models usually need to deal with; the model problem is when NLP models are trained on large-scale datasets that may have problems such as bias and noise, which affect the predictive power and accuracy of the model. Predictability is something we need to pay attention to because it can have a series of knock-on effects that lead to results of natural language processing tasks that are far from the initial task requirements as well as more unpredictable and immeasurable consequences.

Natural language processing techniques may also pose a threat to people's privacy. In speech recognition and semantic analysis, NLP technologies require the collection of speech and text data from individuals. This data may contain sensitive information such as personal identity, health status and financial information. If this data is leaked or misused, it will pose a threat to the privacy and security of individuals. Therefore, researchers need to take steps to ensure the right to privacy of personal data. The right to privacy is the most basic human right, and only by maximizing the protection of individual privacy can artificial intelligence natural language processing develop on a path that does not deviate from its ethical meaning.

Unclear liability regimes for AI natural language processing technologies may lead to undesirable consequences and can raise ethical liability issues. Standards of responsible practice need to be established to reduce potential adverse effects and to hold developers and users of algorithms accountable. Increased regulation and review are needed to pay attention to liability issues from the beginning of the design of language models. The application of natural language processing to decision making should pay more attention to these issues and ensure the reasonableness and fairness of the results of evaluation and decision making from the technical level.

Language models of natural language processing techniques may be biased because their training data are often obtained from human-written texts. Therefore, if biases are present in the training data, the language models may also exhibit the same biases. These biases may affect the impartiality and credibility of NLP techniques. Therefore, researchers need to understand these biases scientifically from an ethical perspective and think about the causes of the biases to correct them at the root cause to ensure impartiality, credibility, and the rights of the recipients.

In summary, there are certainly many ethical issues in artificial intelligence natural language processing, but we must take a two-pronged approach from the subjective perspective of technology creation and the objective perspective of social criticism and must pay attention to ethical issues in the development process of natural language processing itself, such as algorithm, model building, and technology integration, as well as in social evaluation and social impact to ensure the benign development of natural language processing. The ethical issues brought about by technological development are worthy of attention, and measures must be taken to mitigate the impact of these issues while ensuring that the development of NLP technology can maximize the benefits to human beings. In addressing these ethical issues, rigorous quality control and review mechanisms, the use of more diverse data for training, and fairness assessment methods are needed to ensure that NLP technologies can maximize human well-being while following ethical and moral guidelines.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Behera, R. K., Bala, P. K., Rana, N. P., & Irani, Z. (2023). Responsible natural language processing: A principlist framework for social benefits. *Technological Forecasting and Social Change, 188*, Article 122306. https://doi.org/10.1016/j.techfore.2022.122306

[2] Garrido-Muñoz, I., Montejo-Ráez, A., Martínez-Santiago, F., & Ureña-López, L. A. (2021). A survey on bias in deep NLP. *Applied Sciences, 11*(7), Article 3184.

[3] Hovy, D., & Prabhumoye, S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass, 15*(8), Article e12432. https://doi.org/https://doi.org/10.1111/lnc3.12432

[4] Jin, Z., & Mihalcea, R. (2022). Natural language processing for policymaking. In E. Bertoni, M. Fontana, L. Gabrielli, S. Signorelli, & M. Vespe (Eds.), *Handbook of computational social science for policy* (pp. 141-162). Springer.

[5] Königs, P. (2022). Artificial intelligence and responsibility gaps: What is the problem? *Ethics and Information Technology, 24*(3), Article 36. https://doi.org/10.1007/s10676-022-09643-0

[6] Kreutzer, T., Vinck, P., Pham, P. N., An, A., Appel, L., DeLuca, E., Tang, G., Alzghool, M., Hachhethu, K., Morris, B., Walton-Ellery, S. L., Crowley, J., & Orbinski, J. (2020). Improving humanitarian needs assessments through natural language processing. *IBM Journal of Research and Development, 64*(1/2), 9:1-9:14. https://doi.org/10.1147/JRD.2019.2947014

[7] Liu, X., Xie, L., Wang, Y., Zou, J., Xiong, J., Ying, Z., & Vasilakos, A. V. (2021). Privacy and security issues in deep learning: A survey. *IEEE Access, 9*, 4566-4593. https://doi.org/10.1109/ACCESS.2020.3045078

[8] Stanczak, K., & Augenstein, I. (2021). A survey on gender bias in natural language processing. *arXiv preprint*, arXiv:2112.14168.

[9] Tang, C. (2022). Privacy protection dilemma and improved algorithm construction based on deep learning in the era of artificial intelligence. *Security and Communication Networks, 2022*, Article 8711962. https://doi.org/10.1155/2022/8711962

[10] Xiao, Y., & Wang, W. Y. (2019). Quantifying uncertainties in natural language processing tasks. *Proceedings of the AAAI conference on artificial intelligence, 33*(1), 7322-7329. https://doi.org/10.1609/aaai.v33i01.33017322