DSI DATA SCIENCE AND INFORMETRICS
VOL.3, NO.4, NOV. 2023

RESEARCH ARTICLE

Uncovering the influence of ChatGPT´s prompts on traffic safety information using text mining approach

Boniphace Kutela[a*], Kelvin J. Msechu[b], Norris Novat[c], Emmanuel Kidando[d], Angela E. Kitali[e]

a. Texas A&M Transportation Institute, Houston, USA
b. Atkins North America Inc., Novi, USA
c. Iteris Inc., Fairfax, USA
d. Department of Civil and Environmental Engineering, Cleveland State University, Cleveland, USA
e. School of Engineering and Technology, University of Washington Tacoma, Tacoma, USA

ABSTRACT
ChatGPT has emerged as a promising advanced large language model that needs prompt to gain information. However, designing a good prompt is not an easy task for many end-users. Therefore, this study intends to determine the amount of information gained because of varied amounts of information in the prompt. This study used two types of prompts, initial and improved, to query the introduction sections of 327 highly cited articles on traffic safety. The queried introduction sections were then matched with the corresponding human-written introduction sections from the same articles. Similarity tests and text network analysis were used to understand the level of similarities and the content of ChatGPT-generated and human-written introductions. The findings indicate the improved prompts, which have the addition of generic persona and information about the citations and references, changed the ChatGPT's output insignificantly. While the perfect similar contents are supposed to have a 1.0 similarity score, the initial and improved prompt's introduction materials have average similarity scores of 0.56 and 0.54, respectively. Further, the content analysis revealed that themes such as *statistics, trends, safety measures*, and *safety technologies* are more likely to have high similarity scores, irrespective of the amount of information provided in the prompt. On the other hand, themes such as *human behavior, policy and regulations, public perception,* and *emerging technologies* require a detailed level of information in their prompt to produce materials that are close to human-written materials. The prompt engineers can use the findings to evaluate their outputs and improve their prompting skills.

KEYWORDS
ChatGPT; Artificial intelligence; Prompt; Traffic safety; Text mining

1 Introduction

The use of Artificial Intelligence (AI) has continuously increased and penetrated various sectors that involve human interactions, operations, and well-being. AI has aided in acceler-

* Corresponding Author: b–kutela@tti.tamu.edu

ating technological advancement, creation, and simplification of jobs, and even simplifying some of the basic human tasks. Good examples of common day-to-day use of AI are through voice control devices like Siri, Google Assistant, and Amazon Alexa, as well as the use of smartwatches and smartphones to monitor and track various activities such as fitness and health data.

The extensive use and advancement of AI have seen a tremendous surge, particularly in language modeling. Recently, OpenAI launched an interactive online chatbot (ChatGPT) that generates responses based on the prompt provided by the user (Liu et al., 2023; Noever & McKee, 2023a; Perlman, 2022). This advancement is enabled by a Generative Pre-trained Transformer (GPT) trained using a dataset of over 570 GB of text data with 175 billion parameters (OpenAI, 2020). Due to the vast amount of training data used, ChatGPT can perform several human tasks, which led to gaining attention and popularity in various fields, including engineering, law, medicine, computer engineering, and education (Bommarito & Katz, 2022; Perlman, 2022)

ChatGPT interacts with a user through a prompt. The prompt is a textual message or statement the user provides to initiate a conversation or a specific question (Schmid, 2023). The prompt can be in the form of a sentence or a phrase and can be as simple or complex as the user desires. For example, a user may enter the prompt, "Can you tell me what the weather will be like tomorrow?" or "Describe the severe weather event in the United States in 2020". Furthermore, the prompts used in ChatGPT are diverse and can cover a wide range of topics, from simple queries to complex statements. The system uses advanced natural language processing techniques to analyze and understand the meaning of the prompt and generate an appropriate response, whether that response is informative, supportive, or entertaining.

With the variety of ways of preparing prompts to retrieve information from ChatGPT, the question arises on to what extent does altering the inputs in the prompts changes the output. Specifically, how the inclusion of persona affects the retrieved information from ChatGPT. The examination of the impacts of the changes in the prompt input is necessary to understand the variation of the information one person can obtain compared to another person, if they use slightly different prompts. To examine such changes is relatively difficult as it requires the ground truth information to compare with ChatGPT output. Thus, this study utilized the published introduction sections of the research papers as the ground truth data and varied prompts to obtain different outputs from ChatGPT. The next section of the paper presents the related works followed by the methodology. The results and discussion are then presented, followed by the conclusion and study limitations.

2 Related Work

ChatGPT has emerged as one of the promising AI tools for various tasks in research, academia, and day-to-day purposes. In academia, various researchers have examined its capabilities to pass various exams (Chalkidis, 2023; Newton & Xiromeriti, 2023). Several researchers have evaluated its ability to interpret some difficult scenarios that could otherwise take large efforts. Such attempts include legal services, traffic safety, drug discovery and medical procedures, among others (Chalkidis, 2023; Sharma & Thakur, 2023; Zheng et al., 2023)

The application of ChatGPT in scientific writing has gained the interest of various stakeholders (Gao et al., 2022; Macdonald et al., 2023; Newton & Xiromeriti, 2023; Zheng & Zhan, 2023). Evidently, AI is changing the way we approach scientific writing. Recently, there has been growing interest in the role of ChatGPT in scientific writing. This interest can be attributed to the rapid advancements in AI and its potential applications in various academic fields. Zheng & Zhan (2023) studied this topic, highlighting the challenges and ethical dilemmas that researchers might face when using ChatGPT for academic purposes. They emphasized the importance of understanding the underlying mechanisms of ChatGPT and the potential biases it might introduce into scientific content. Their insights are crucial for the current study as we also aim to understand how effective ChatGPT is in generating scientific content. On a similar note, Dashti et al. (2023) conducted a comprehensive study on the reliability of AI tools, specifically focusing on ChatGPT, in the realm of scientific writing. Their research revealed that while AI tools like ChatGPT have the potential to revolutionize scientific writing by automating certain tasks, they are not without flaws. The generated content, although coherent, might lack the depth and reason of human-written content. Hence, their findings suggest that while these tools can be instrumental in aiding researchers, they should be used with caution and with a clear understanding of their limitations.

The quality of abstracts generated by ChatGPT has also been a topic of discussion. Gao et al. (2023) compared abstracts produced by ChatGPT with real ones. They found that ChatGPT can create clear abstracts, but they might lack depth. This finding is crucial for our research as we also use machine learning to assess ChatGPT's content.

The style of writing in academic papers is not just a matter of formality; it plays a pivotal role in conveying complex ideas clearly and effectively. Lu et al. (2019) investigated the nature of linguistic complexity in scientific writing and its subsequent influence on how a paper is received by its audience. Their findings underscored that the manner in which information is presented can significantly affect comprehension, engagement, and the overall impact of a paper. For instance, a well-structured and clearly written paper can facilitate better understanding, even for readers who might not be experts in the specific field. On the other hand, overly complex or convoluted writing can alienate readers, potentially obscuring the core message of the research. This balance between clarity and depth is crucial in ensuring that a paper reaches and resonates with its intended audience. In the context of the current study, in capabilities of ChatGPT in generating scientific content, understanding the benchmarks of good academic writing is paramount. If ChatGPT is to be utilized as a tool for academic writing, it's essential to ensure that its outputs align with the established standards of clarity, coherence, and complexity that studies like that of Lu et al. (2019) have highlighted.

Other studies, like those by Ali & Djalilian (2023), have also highlighted the potential and pitfalls of using ChatGPT in scientific writing. They emphasize the need for caution and the ethical considerations that come with using AI tools in academia. There's also concern about AI tools generating misleading scientific papers. Májovský et al. (2023) explored this, showing the risks of relying too heavily on AI for scientific research. The ethics of using AI in academic writing is a concern in scientific writing topics. Balat & Bahsi (2023) discussed whether AI tools, no matter how advanced, should be credited as authors. They believe that these tools can't be held accountable for what they write.

The success of the response generated by ChatGPT depends on the quality and clarity of the prompt input by the user (McCue, 2023). Thus, the basic prompt should be clear, relevant, complete, and concise (Slater, 2023). This is to say the information provided to ChatGPT should be relevant to the chatbot's domain or area of expertise and contains enough information for ChatGPT to understand the user's intention. Further, it should be concise, meaning it should not be too long or wordy as it may be difficult for ChatGPT to understand, and it should be courteous and use appropriate language (Cooper, 2023; Schmid, 2023; Slater, 2023). Researchers agree that a well-drafted prompt results in better and more relevant results from ChatGPT. To improve prompt results, the context of the information sought is important. Such context may involve the use of persona, which is impersonating a certain person who is well familiar with the topic of interest (Kocaballi, 2023; Noever & McKee, 2023b; Probert, 2023). Adding a persona intends to improve results through personalization, better targeting, and improved efficiency (Probert, 2023).

Although various studies have utilized prompts and hypothesize that a well-drafted prompt produces a better output from ChatGPT, no scientific or statistical evidence has been provided. For instance, it is unknown to what extent the ChatGPT outputs improve with adding the persona. Further, creating a well-drafted prompt requires more words, which is against one of the basic principles of the prompts : they should not be too long and wordy. Thus, this study intends to explore the extent to which a well-drafted prompt improves the information obtained from ChatGPT. Further, the study explores not only the statistical evidence but also the content of the text resulting from the varying levels of prompt descriptions. In this context, the text obtained through varying inputs in the prompt is compared to the human-written texts on the same topic. The results of this study will provide valuable insights into the capabilities of the ChatGPT's prompts in retrieving information given varying levels of inputs. The study will have important implications for the future of scientific writing and the integration of AI in academic research and public use.

The rest of the paper is organized as follows. The next section of the paper presents the methodology, which outlines the data description and the analytical methods used in the study. This section is followed by the results and discussion section, where the study findings are presented and analyzed in detail. Finally, the conclusion section provides a summary of the key findings, along with recommendations for future research is presented.

3 Methodology

As described earlier, this study intends to explore whether prompt clarity improves the ChatGPT's content by providing materials as close as human-written materials. In addition, the study intends to understand the contents of the ChatGPT materials resulting from different prompts. To do so, the similarity score analysis and content analysis are applied. This section presents the methodology, divided into two main sections: data description and analytical methods. The analytical methods include document similarity analysis, which evaluates how similar the documents are, and text network analysis, which describes the documents' content.

3.1 Data Acquisition

In this study, the authors utilized the introduction section of published papers as the comparison data to the ChatGPT data. Although some of the previous studies have utilized abstracts to evaluate the capabilities of ChatGPT (Gao et al., 2022; Kutela, Msechu, Das, et al., 2023; Macdonald et al., 2023), authors hypothesized that it is comparatively easy even for a person to prepare an introduction than abstract given a heading or title of the paper. Thus, the introduction sections were utilized in this study.

The authors used papers focused on traffic safety published in various journals. The papers were retrieved from the Web of Science database (Clarivate, 2023) using the keywords "traffic safety", "traffic crash", "transportation safety", "vehicle accident", and "traffic accident" in the abstracts. These keywords were based on the authors' understanding of the keywords used in the traffic safety studies since authors are experienced researchers in the traffic safety field. The papers were ranked based on the number of citations, whereby only papers with a minimum of 30 citations were included for further analysis.

The paper's title was used to create the ChatGPT-generated text. Two prompts, the initial prompt and the improved prompt, were prepared. The key difference between the two prompts is that the initial prompt had basic information, while the improved prompt had an additional description of generic persona and the information for citations, references, and statistics. The initial prompt reads as follows.

> *I want you to develop an introduction section of a manuscript for publication. I will give you a number of titles then I want you to give me the introduction section of the paper. The first title is "Title of the paper".*

On the other hand, the improved prompt reads as follows.

> *I want you to develop an introduction section of a manuscript for publication. I will give you a number of titles then I want you to give me the introduction section of the paper. You need to adopt a persona of a highly skilled writer in traffic safety. In your writeup, include the actual citations, actual references, and actual traffic safety statistics. The first title is "Title of the paper".*

The ChatGPT-generated introduction s were established by using the 327 original manuscript titles as inputs. These titles were provided to the ChatGPT prompt. To evaluate the output of the ChatGPT-generated introductions, each was paired with its human-written equivalent in an Excel spreadsheet for a comprehensive analysis.

3.2 Analytical Methods

This section presents a discussion of the analytical methods. Two approaches are presented: document similarity analysis and text network analysis. The document similarity analysis presents the similarities between the human written and ChatGPT-based materials whereby the two levels of prompts are assessed against the human written materials and across themselves. On the other hand, the text network analysis uses keywords and pairs of co-occurred keywords to show the differences in the content of the materials.

3.2.1 Document Similarity Analysis

Document similarity analysis is a widely used methodology for measuring the degree of similarity between different documents. Typically, researchers and the general public alike assume that documents are similar if they exhibit semantic closeness and share similar concepts or themes.

The present study focuses on the evaluation of document similarity analysis by examining the degree of similarity between the introduction sections of the paper generated by ChatGPT and compared to those written by a human. Specifically, the bag-of-words representation is applied to compute the similarity between two documents using the Cosine similarity with Latent Semantic Analysis (LSA) approach. The similarity score between the human-written introduction section and the ChatGPT-generated introduction is computed using Cosine similarity score (Han et al., 2012; Lahitani et al., 2016) presented in Equation 1.

$$Similarity\,(doc_1, doc_2) = Cos\theta = \frac{doc_1\, doc_2}{|doc_1||doc_2|} \tag{1}$$

Whereby doc_1 is the human-written introduction of the paper and doc_2 is the ChatGPT-generated introduction of the same paper.

To perform cosine similarity, the score of the documents are first converted to a vectorized form of representation (Lahitani et al., 2016). The vectors are then used to determine the cosine relationship between them using statistical packages such as text2vec (Selivanov, 2022). If the cosine score between the two vectors is 1 or closer, the documents are said to be similar, otherwise dissimilar (Han et al., 2012; Lahitani et al., 2016; Zach, 2020). Thus, if the two documents generated by adding a general persona in the prompt differ significantly, the implication is that the prompt used to create the document has influenced the outcome. The vise versa is true if the documents do not differ significantly.

3.2.2 Text Network Analysis

Text Network Analysis (TNA) has proven to be a valuable tool in a wide range of fields, including literature and linguistics (Hunter, 2014) and traffic safety and operations (Kutela, Das, et al., 2021; Kutela, Kadeha, Magehema, et al., 2023; Kutela & Teng, 2021) among others. By utilizing nodes and edges, TNA can establish relationships between keywords within a given corpus, resulting in a network representation of the underlying structure (see **Figure 1**). TNA's greatest strength is its ability to visually represent the relationships among keywords, providing a clear and intuitive view of the connections between various elements (Jiang et al. , 2020; B. Kutela et al., 2021; Boniphace Kutela et al., 2021; Paranyushkin, 2011). In a TNA network, the size of the nodes corresponds to the frequency of the keywords within the corpus, while the size of the edges represents the degree of co-occurrence between them.

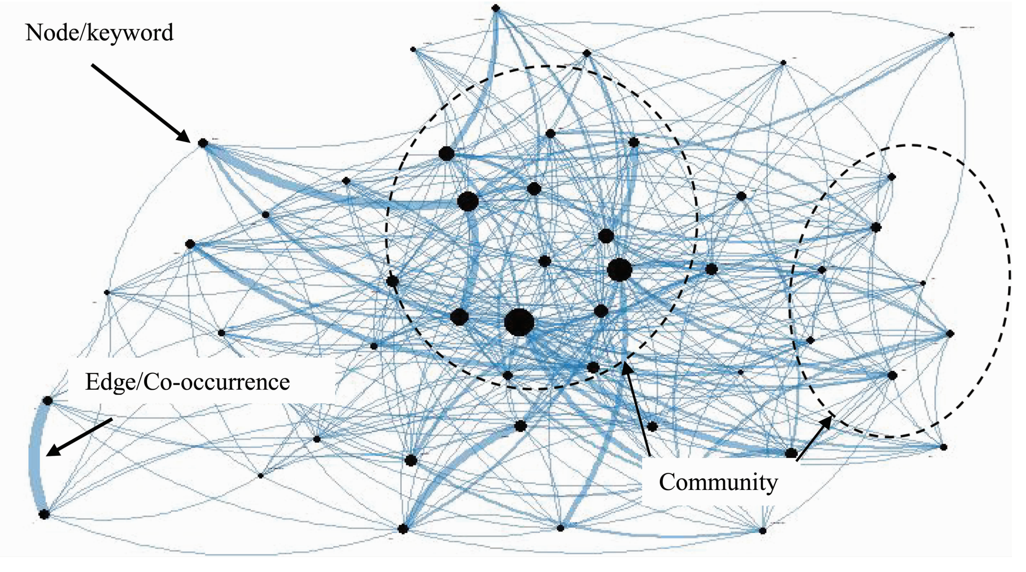


Figure 1 A skeleton of the text network

In the TNA analysis process, several critical steps are involved. The first step is data normalization, where the raw unstructured data is converted into structured data, and all symbols are removed while converting all text to lowercase (Kutela, Magehema, et al., 2022; Kutela, Novat, et al., 2022; Paranyushkin, 2011). The normalized data is then used to create a matrix of keywords with their corresponding frequencies of occurrence. The matrix is then visualized as a network of nodes, with the size of each node representing the frequency of the corresponding keyword. Comparative analysis is performed using various metrics, with this study utilizing document and collocated frequency to compare the introductions generated by humans and ChatGPT. Document frequency indicates the number of documents that contain a particular keyword, while keyword frequency measures how many times the keyword appears in a document (Kutela, Msechu, Kidando, et al., 2023; Kutela, Oscar, Kidando, et al., 2023). Collocation frequency determines the number of times keywords appear next to each other and provides a deeper insight into the relationships between keywords. The collocation of keywords in a text network plays a significant role in forming text clusters or communities of keywords, representing a group of keywords clustered together in the network (Kutela, Combs, et al., 2022; Kutela, Kitali, et al., 2022; Paranyushkin, 2011). A text network may have multiple communities (see **Figure 1**), with each community consisting of closely related and strongly connected keywords (Kutela, Dzinyela, Haule, et al., 2023).

4 Descriptive Summary of the Data

The basis for comparing the data is needed to explore the capability of the ChatGPT prompt. Initially, a list of 525 papers was retried from the Web of Science database. The titles of the papers were then checked, and 102 papers that were not about traffic safety were removed. Also, 96 papers from journals that were not accessible through the authors' library were taken out. Finally, 327 papers from 102 journals were selected for further study, with the majority coming from the *Journal of Accident Analysis and Prevention, the Journal of Safety Research*, and *Analytic Methods in Accident Research* as indicated in **Figure 2**. In addition to the papers presented in **Figure 2**, other 75 papers were from 75 different journals. Additionally, **Figure 3** presents the distribution of published papers by years. It can be observed that most of the papers used in this study were published in 2019 and 2020, while the oldest paper was published in 1995.

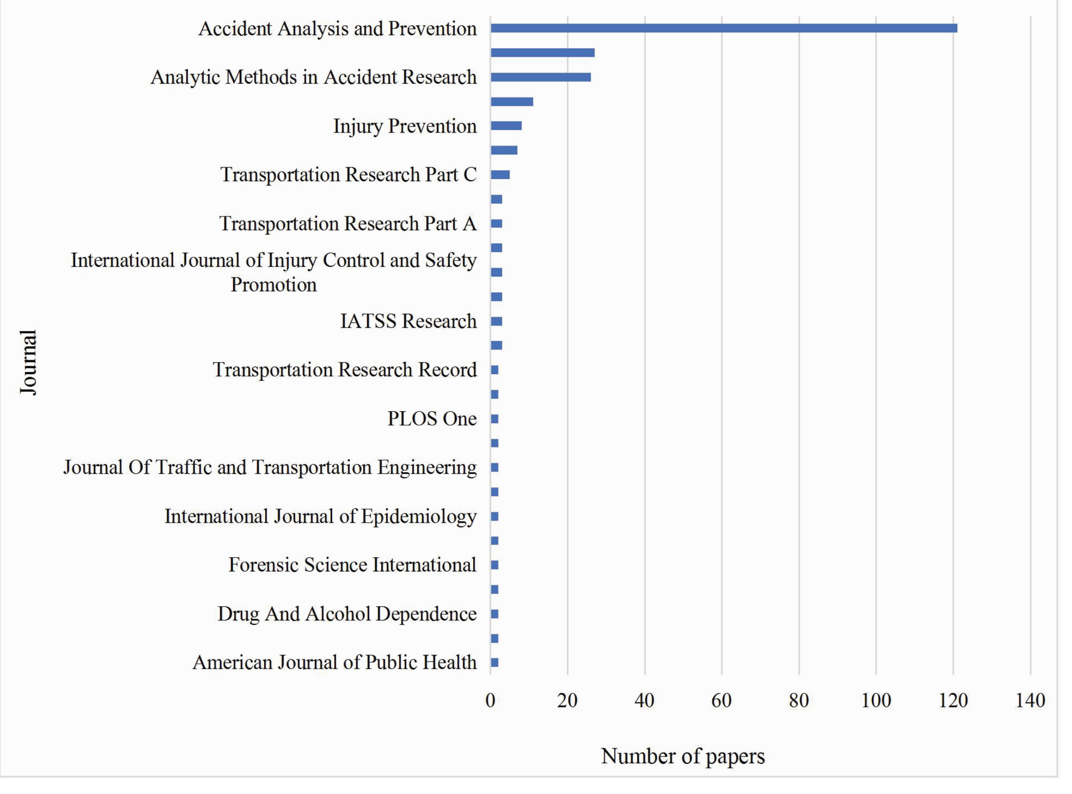


Figure 2 Distribution of the number of papers per journal

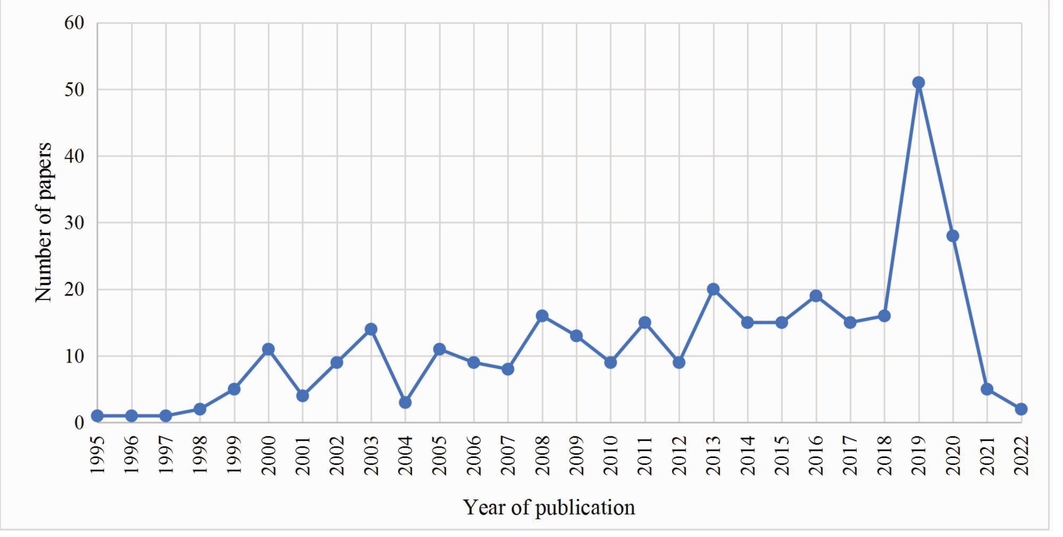


Figure 3 Distribution of the number of papers by years

5 Results and Discussions

This section presents the results and discussion of this study. It covers the similarity scores, which show the magnitude of similarities between human-written and ChatGPT-generated materials and the text network results that depict the content of the materials.

5.1 Similarity Scores Results

Table 1 presents the similarity score results for the three comparisons: improved prompt against human written, initial prompt against human written, and improved prompt against the initial prompt. The similarity scores range from 0.0633 to 0.9638, with an average score of 0.5567 for improved prompt vs. human written, 0.5387 for initial prompt vs. human written, and 0.8158 for improved prompt vs. initial prompt. The similarity average scores result from comparing prompt-generated introductions to human-written introductions suggests both improved and initial prompts average similarity scores (0.5567, 0.5387, respectively) have low similarity to human-written texts. However, the gain of similarity score, about 2%, is relatively small. When comparing the improved and initial prompts, the improved prompt produced more similar text, with an average similarity score of 0.8158. This suggests that the level of information in the prompt had no significant impact on the output generated by ChatGPT. Thus, to obtain more improved similarities scores, users may utilize specific personas.

Table 1 Similarity score statistics

Comparison	Similarity Score Statistics					
	Minimum	1st Quartile	Median	Mean	3rd Quartile	Maximum
Improved prompt vs. Human written	0.0040	0.3914	0.5910	0.5567	0.7523	0.9638
Initial prompt vs. Human written	0.0633	0.3816	0.5605	0.5387	0.7286	0.9784
Improved prompt vs. Initial prompt	0.0005	0.7874	0.8678	0.8158	0.9287	0.9905

Overall, the similarity score statistics provide insight into the effectiveness of using ChatGPT to generate text in response to prompts. The methodology and findings of this study can be used to inform future research on optimizing the use of AI language models in various fields. It is worth noting that while the use of the initial prompt may result in lower similarity scores, it still has its value in certain contexts. The initial prompt may be appropriate when the goal is to obtain a more diverse range of responses or to encourage the model to generate more creative or unexpected outputs. In addition, the initial prompt may be more useful in cases where the available information is limited or where the user is uncertain about the exact phrasing or structure of the prompt. Therefore, the decision to use the initial/weak prompt should be based on the specific goals and requirements of the task and should not be dismissed outright as inferior to the improved prompt.

5.2 Text Networks Results

This section presents the text network results for human-generated introductions and ChatGPT-generated introductions.

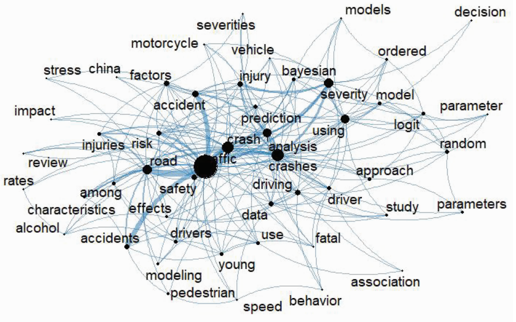
5.2.1 Text Network of the Titles

The text in **Figure 4** displays networks comparing the ChatGPT-generated introductions with high and low similarity scores for both improved and initial prompts. These networks exhibit clusters of nodes with high connectivity based on the words they contain. **Figure 4(a)** shows the network for improved prompt and high similarity score, with the keyword *traffic* at the center, and has dense edges with nodes such as *accidents, road, crash, injuries*, and *risk*. The clusters are related to Bayesian modeling and its parameters.

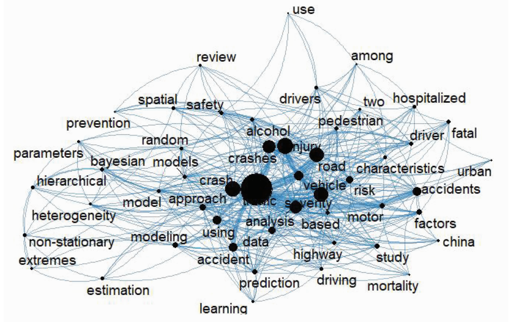
In contrast, **Figure 4(b)** displays the network for improved prompt and low similarity score, with clusters centered on *traffic* and *parameters*. The traffic cluster nodes are *crashes, approach, severity, road, safety*, and *aim*, while the parameter clusters are related to Bayesian modeling and its various components. For the initial prompt and high similarity score net work(**Figure 4(c)**), *traffic* is still the center, but other larger nodes, such as *crashes, road, severity, factors*, and *vehicle*, have dense edges with the *traffic* node. The cluster is related to Bayesian modeling and its parameters. Lastly, **Figure 4(d)** presents the network for initial prompt and low similarity score, with dense nodes on *traffic, analysis, data, severity, road,* and *crash*. It also has smaller clusters related to Bayesian modeling and its parameters. However, the overall connectivity is less compared to the other networks.

Overall, the text networks in **Figure 4** demonstrate that the clusters of nodes in ChatGPT-generated introductions are highly dependent on the strength and similarity of the prompt. Both improved prompts and high similarity scores and initial prompts and low similarity scores result in diverse and less cohesive clusters.

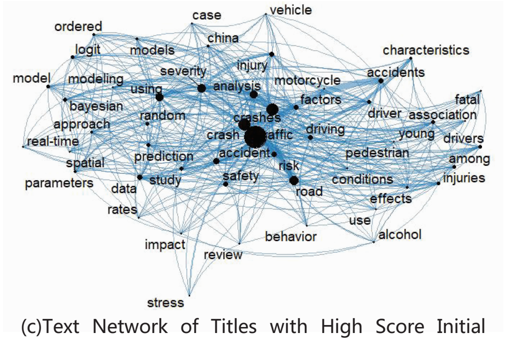
Table 2 provides a comparison of high similarity scores and low similarity scores between the improved prompt and the initial prompt. The feature, frequency, and document frequency of each prompt are presented in the table. The high similarity scores suggest that there are several common features that are likely to result in a similar output, regardless of the level of information provided in the prompt. For instance, the features *traffic, crash, accident, road, severe*, and *use* are highly frequent in both the improved and initial prompts and have a high document frequency as well. Other common features include *analysis, driver, model, safety, risk*, and *study*, which are found in both prompts with a high frequency and document frequency. These similarities indicate that certain features are likely to produce consistent output from the AI language model, regardless of the level of detail provided in the prompt.



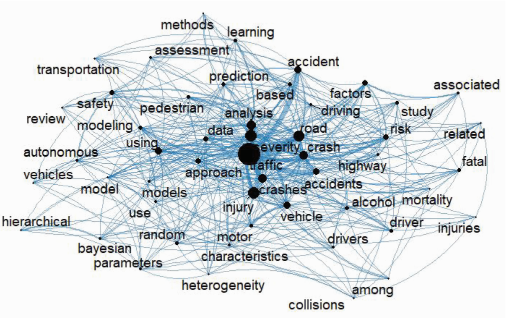
(a)Text Network of Titles with High Score Improved prompt



(b)Text Network of Titles with Low Score Improved prompt



(c)Text Network of Titles with High Score Initial prompt



(d)Text Network of Titles with Low Score Initial prompt

Figure 4 The text network of the titles

Table 2 Comparison of high similarity scores and low similarity scores between the improved prompt and the initial prompt

High Similarity Scores						Low Similarity Scores					
Improved prompt			Initial prompt			Improved prompt			Initial prompt		
feature	**freq**	**docfreq**	**feature**	**freq**	**docfreq**	**feature**	**freq**	**docfreq**	**feature**	**freq**	**docfreq**
traffic	98	94	traffic	96	92	traffic	63	63	traffic	65	65
crash	85	82	crash	88	85	crash	53	53	crash	50	50
accident	54	53	accident	52	51	accident	35	35	accident	37	37
road	43	42	road	42	41	road	30	28	injuries	32	32
sever	37	37	sever	32	32	injuries	28	28	sever	32	32
use	37	36	use	32	32	sever	27	27	road	31	29
injuries	33	33	model	34	31	model	28	26	analysis	26	26
analysis	33	33	analysis	30	30	analysis	23	23	model	26	25
driver	33	31	driver	30	29	use	17	17	use	22	21
model	32	30	injuries	29	29	vehicle	16	16	vehicle	21	21
safeties	27	26	safeties	23	23	data	14	14	risk	15	15
risk	25	25	risk	23	23	risk	13	13	driver	16	14
factor	23	23	data	24	22	driver	13	12	approach	14	14
studies	23	23	studies	22	22	approach	12	12	safeties	15	13
data	23	21	factor	21	21	studies	12	11	pedestrian	13	13
effect	21	21	effect	20	20	predict	11	11	data	13	13
drive	23	20	drive	22	19	pedestrian	11	11	factor	13	13
predict	14	14	predict	15	15	factor	11	11	studies	13	12
fatal	14	14	bayesian	12	12	safeties	11	10	predict	10	10
behavior	13	13	young	12	12	spatial	8	8	fatal	10	10

Note: Docfreq= document frequency

The high similarity scores in the table suggest that the improved and initial prompts are generating similar results, particularly when it comes to the frequency and document frequency of certain terms. Both prompts produced the same top five most frequent terms (*traffic, crash, accident, road*, and *severe*), indicating that the prompts are generally effective at extracting the most relevant information. Additionally, both prompts have similar frequencies for terms related to driver behavior and risk factors, indicating that these themes are salient in the introduction sections of highly cited articles on traffic safety.

However, it is important to note that the initial prompt generated lower frequencies for some of the terms compared to the improved prompt, such as *injury, analysis*, and *model*. This could suggest that the improved prompt being well drafted, can provide further emphasis/elaborations on keywords related to the specified topic. Furthermore, the initial prompt generated some terms (such as *bayesian* and *young*) that were not present in the improved prompt. While these terms may be relevant to some articles on traffic safety, their lower frequency suggests that they are not as prominent as the terms captured by the improved prompt. Overall, while the high similarity scores between the improved and initial prompts are promising, further analysis is needed to determine the impact of these differences in frequencies on the quality of the generated output.

The low similarity scores in the table show that the frequency and document frequency of features in the improved prompt and initial prompt are quite different, indicating that the prompts are not as similar to each other as compared to the high similarity scores table. For instance, the word *injury* and *severe* have higher frequencies in the initial prompt as compared to the improved prompt, while the words *model* and *analysis* have higher frequencies in the improved prompt. Furthermore, some features, such as *spatial* and *fatal*, only appear in the initial prompt and not in the improved prompt, indicating that the prompts have different focuses and content.

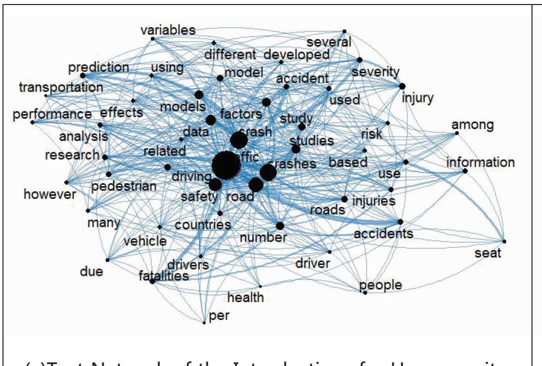
Moreover, the similarities and differences in term frequencies between the improved and initial prompts suggest that both approaches have their strengths and weaknesses in extracting relevant information from highly cited articles on traffic safety. While the high similarity scores are promising, it is important to consider the potential impact of the differences in frequencies on the quality of the output. Ultimately, the selection of a prompt will depend on the specific research question and the level of detail required in the generated output. The difference in the frequencies of the features in the improved prompt and initial prompt suggests that the type of prompt used can have a slight impact on the results of the study. Therefore, researchers should be careful when choosing prompts for their studies and ensure that the prompts are relevant to the research questions and objectives. Additionally, **Table 2** shows that even though the prompts have similar topics, their underlying meanings and connotations can be quite different. As such, it is crucial for researchers to carefully consider the language and wording used in their prompts to ensure that they accurately capture the intended meaning and do not introduce unintended biases into the study.

5.2.2 Text Network for Human Written and Improved prompt–generated Introductions.

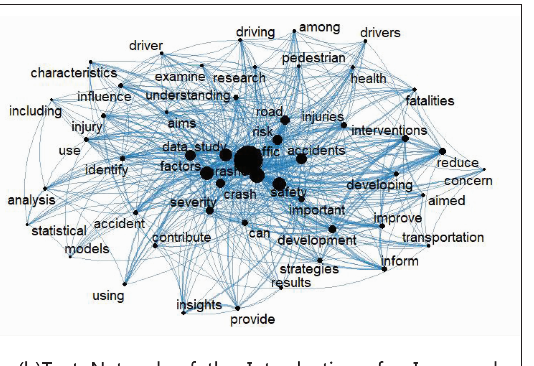
The results presented in **Figure 5** show four different networks of Introductions sections, with both high and low similarity scores, for both ChatGPT-generated and human-written texts. Interestingly, all networks are centered on the keyword *traffic.* Specifically, the network for human-written introduction sections with high similarity scores (**Figure 5(a)**) shows dense edges with nodes such as *fatalities, road, crash*, and *accidents.* The text network connected edges such *as traffic-crash-factors, traffic-road-accidents*, *traffic-crash-data*, and *traffic-fatality* suggest that the human-generated introductions focus more on the causes and outcomes of traffic crashes. This is further emphasized with keywords such as *severity, injuries*, and *risk* present in the network.

On the other hand, the network for ChatGPT-generated introduction sections with high similarity scores (**Figure 5(b)**) shows dense edges with nodes such as *study, accidents, risk, factors, safety*, and *crash*. The dense text network edges from the ChatGPT-generated introduction are similar to the human-generated network. However, the ChatGPT network contains keywords such as *improve, inform, reduce, understanding,* and *interventions,* to mention a few that focus on suggestions for actions to reduce traffic incidents. These findings suggest that human writers and ChatGPT focus on different aspects of traffic safety. In contrast, ChatGPT-generated keywords tend to focus on generic words related to traffic safety solutions. Overall, the networks presented in Figure 3 visually represent the common themes and topics discussed in the introduction sections on traffic-related research and highlight the potential for using text network analysis to better understand the structure and content of scientific literature.

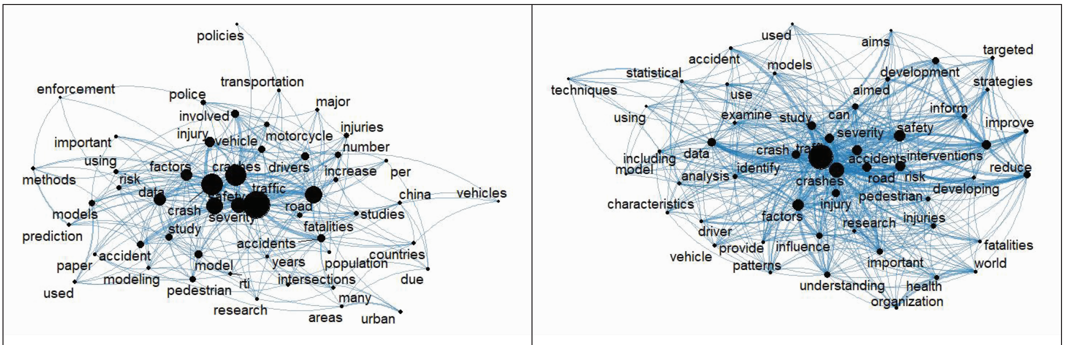
According to the network for human written-introduction sections with low similarity scores (**Figure 5(c)**), the key node *traffic* has relatively larger nodes with dense edges, including *crashes, road, severity, factors,* and *vehicle*. These nodes are all related to the topic of road traffic accidents and their causes. Additionally, the cluster of nodes, including *prediction, models, method, paper, used,* and *modelling* indicates a focus on research methodology and data analysis. The cluster of nodes, including *intersections, urban*, and *areas* suggests a focus on the contextual factors related to road safety. These introductions may be discussing road safety measures that are specific to urban areas or intersections, which may be associated with higher rates of accidents. In contrast, the network for ChatGPT-generated introduction sections with a low similarity score (**Figure 5(d)**) has a key node *traffic* with dense edges with nodes including *study, factors, data, influence, safety*, and *severity*. These nodes suggest a focus on the factors contributing to road traffic accidents and their impact on safety. The cluster of nodes, including *reduce, improve, inform, interventions, strategies*, and *targeted*, indicate a focus on potential interventions to reduce the incidence of accidents and improve road safety.



(a)Text Network of the Introductions for Human-written (High Similarity)



(b)Text Network of the Introductions for Improved prompt (High Similarity)



(c)Text Network the Introductions for Human-written (Low similarity)

(d)Text Network of the Introductions for Improved prompt (Low similarity)

Figure 5 The text network of the introductions for improved prompt

Table 3 Collocation frequency of high similarity scores and low similarity scores between the human written title and the improved prompt title.

Rank	High Similarity Scores			
	Human–written		Improved prompt	
	collocation	count	collocation	count
1	crash prediction models	13	provide important insights	66
2	road traffic crashes	13	can inform development	59
3	world health organization	8	world health organization	58
4	road traffic injuries	6	interventions policies aimed	49
5	road traffic accident	6	road traffic crashes	47
6	leading cause death	6	development targeted interventions	47
7	motorcycle taxi riders	6	public health concern	44
8	use seat belts	5	risk traffic accidents	41
9	traffic flow prediction	5	national highway traffic	40
10	highway traffic safety	4	highway traffic safety	40
	Low Similarity Scores			
Rank	**Human–written**		**Improved prompt**	
	collocation	**count**	**collocation**	**count**
1	road safety strategies	7	world health organization	54
2	road traffic crashes	6	interventions policies aimed	42
3	involved traffic crashes	5	provide important insights	42
4	low– middle–income countries	5	can inform development	41
5	raw count data	4	road traffic crashes	33
6	traffic accident data	3	understanding factors influence	28
7	factors influencing injury	3	development interventions policies	27
8	motor vehicle crashes	3	development targeted interventions	27
9	drivers involved traffic	3	public health concern	26
10	crash severity prediction	3	policies aimed reducing	26

Overall, these results suggest that while the human-written and ChatGPT-generated introductions with low similarity scores both focus on road traffic accidents and their causes, the human-written introductions tend to emphasize research methodology and contextual factors. In contrast, the ChatGPT-generated introductions focus more on potential interventions and strategies to improve road safety. From the results of the networks for high similarity scores, we can see that human-written and ChatGPT-generated introduction sections with improved prompts have nodes strongly connected to the keyword *traffic.* This can be due to the fact that the study titles used are traffic safety related. For human-written sections, we see nodes related to *fatalities, road, crashes, and accidents*. For ChatGPT-generated sections, we see nodes related to *study, accidents, risk, factors, safety,* and *crash*. This suggests that improved prompts can lead to a more focused and relevant discussion of the topic at hand, regardless of whether the text was written by a human or generated by a language model. On the other hand, when we look at the networks for low similarity scores, we see more variation in the nodes and their connections to the keyword *traffic*. This may suggest that weaker prompts can lead to less coherence and relevant discussions.

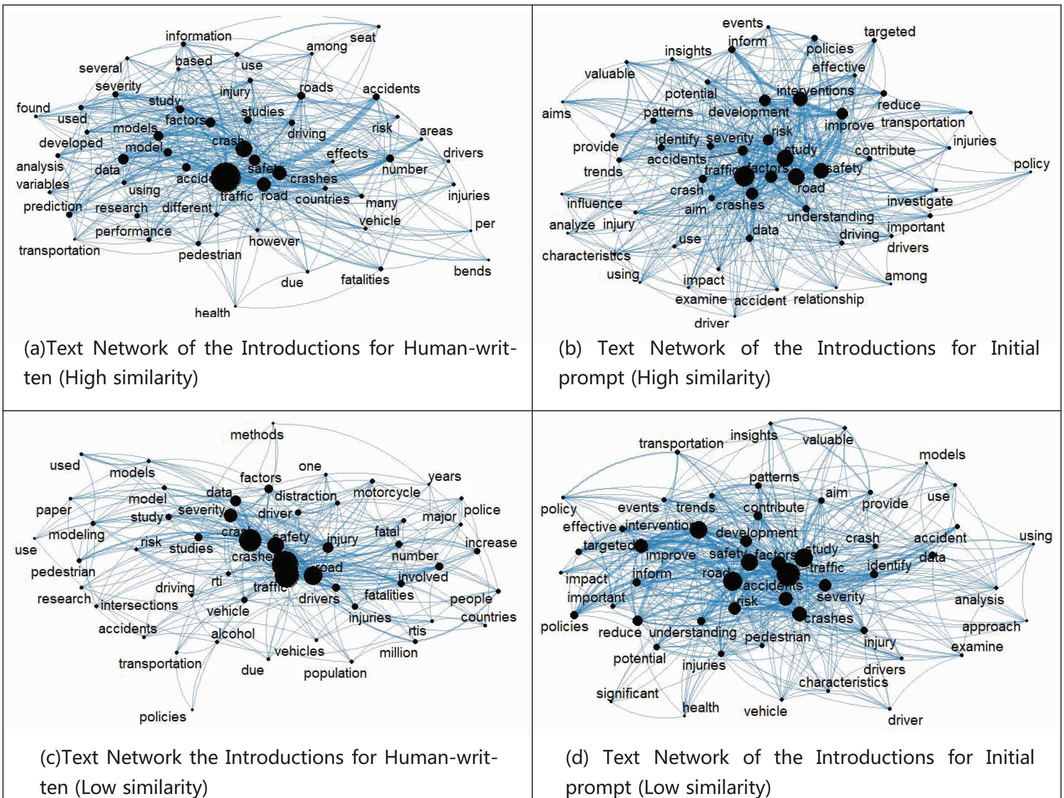
The results of the collocation frequency analysis in(**Figure 5(c)**) show that the key node *traffic* has relatively larger nodes with dense edges, including *crashes, road, severity, factors,* and *vehicle.* These nodes are all related to the topic of road traffic accidents and their causes. Additionally, the cluster of nodes, including *prediction, models, method, paper, used,* and *modelling* indicates a focus on research methodology and data analysis. The cluster of nodes, including *intersections, urban*, and *areas*, suggests a focus on the contextual factors related to road safety. It is possible that these introductions are discussing road safety measures that are specific to urban areas or intersections, which may be associated with higher rates of accidents. In contrast, the network for ChatGPT-generated introduction sections with a low similarity score (**Figure 5(d)**) has a key node *traffic* with dense edges with nodes including *study, factors, data, influence, safety*, and *severity*. These nodes suggest a focus on the factors contributing to road traffic accidents and their impact on safety. The cluster of nodes, including *reduce, improve, inform, interventions, strategies*, and *targeted*, indicates a focus on potential interventions to reduce the incidence of accidents and improve road safety.

Overall, these results suggest that while the human-written and ChatGPT-generated introductions with low similarity scores focus on road traffic accidents and their causes, the human-written introductions tend to emphasize research methodology and contextual factors. In contrast, the ChatGPT-generated introductions focus more on potential interventions and strategies to improve road safety. From the results of the networks for high similarity scores, we can see that human-written and ChatGPT-generated introduction sections with improved prompts have nodes strongly connected to the keyword *traffic*. We see nodes related to *fatalities, road, crashes*, and *accidents* for human-written sections. For ChatGPT-generated sections, we see nodes related to *study, accidents, risk, factors, safety*, and *crash*. This suggests that improved prompts can lead to a more focused and relevant discussion of the topic at hand, regardless of whether the text was written by a human or generated by a language model. On the other hand, when we look at the networks for low similarity scores, we see more variation in the nodes and their connections to the keyword *traffic*. This may suggest that weaker prompts lead to lesser coherence and relevant discussions.

Table 3 above shows notable differences in the use of language between the human-written text and the improved prompt, particularly in the high similarity scores category. The most frequent collocations in the human-written text are *crash prediction models* and *road* *traffic crashes*, each appearing 13 times, while in the improved prompt, the most frequent collocation is *provide important insights*, appearing 66 times. Other frequent collocations in the improved prompt include *can inform development* (59), *world health organization* (58), and *interventions policies aimed* (49). In contrast, the human-written text includes collocations such as *use seat belts* (5) and *motorcycle taxi riders* (6), which are absent or infrequent in the improved prompt.

The low similarity scores table also reveals differences in collocation frequency between the human-written text and improved prompt. The most frequent collocations in the human-written text include *road safety strategies* (7), *road traffic crashes* (6), and *involved traffic crashes* (5), while in the improved prompt, the most frequent collocation is *world health organization* (54). Other frequent collocations in the improved prompt include *interventions policies aimed* (42) and *provide important insights* (42). The counts of collocations in the low similarity scores table are generally lower than those in the high similarity scores table, with the most frequent collocations appearing between 7 and 3 times in the human-written text and between 54 and 26 times in the improved prompt. These results suggest that the language used in the improved prompt uses generic words and differs significantly from that of the human-written text, particularly in terms of frequently occurring collocations, which may affect the outputs generated by the prompt.

5.2.3 Text Network Results for Human Written against Initial prompt–generated Introductions.



(a)Text Network of the Introductions for Human-written (High similarity)

(b) Text Network of the Introductions for Initial prompt (High similarity)

(c)Text Network the Introductions for Human-written (Low similarity)

(d) Text Network of the Introductions for Initial prompt (Low similarity)

Figure 6 The text network of the introductions for initial prompt

The text network in **Figure 6** compares human-written introduction sections with initial prompts given to ChatGPT. The networks with high similarity scores in both human-written and ChatGPT-generated introduction sections (**Figure 6 (a) and (b)**) show dense edges centered on the keyword *traffic*. In the case of human-written introduction sections, the nodes are *traffic, safety, road, crashes*, and *accidents*. In contrast, in the case of ChatGPT-generated introduction sections, the nodes are *traffic, crashes, accidents, severity, road, safety*, and *aim*.

Table 4 Collocation frequency of high similarity scores and low similarity scores between the human written and the initial prompt in comparing introduction sections.

Rank	High Similarity Scores			
	Human–written		Initial prompt	
	collocation	count	collocation	count
1	crash prediction models	13	improve road safety	193
2	road traffic crashes	10	provide valuable insights	85
3	world health organization	8	interventions improve road	82
4	road traffic accident	6	interventions policies improve	57
5	motorcycle taxi riders	6	identify patterns trends	53
6	use seat belts	5	makers transportation planners	52
7	leading cause death	5	policy makers transportation	52
8	traffic flow prediction	5	study potential inform	50
9	road traffic injuries	4	findings study potential	50
10	number bends per	4	potential inform development	50
Rank	**Low Similarity Scores**			
	Human–written		**Initial prompt**	
	collocation	**count**	**collocation**	**count**
1	road traffic crashes	9	improve road safety	128
2	road safety strategies	7	provide valuable insights	50
3	involved traffic crashes	5	interventions improve road	43
4	low– middle–income countries	5	understanding factors contribute	36
5	road traffic injury	4	working improve road	36
6	road traffic injuries	4	road traffic accidents	34
7	number traffic crashes	3	interventions policies improve	32
8	highway traffic safety	3	findings study potential	32
9	drivers involved traffic	3	development targeted interventions	32
10	crash severity prediction	3	study potential inform	31

Moreover, the ChatGPT-generated introduction sections also have a second cluster focused on *development, interventions, inform, policies, effective, targeted, reduce*, and *transportation*. This indicates that while human-written and ChatGPT-generated introduction sections with initial prompts discuss traffic-related topics, there are subtle differences in the clusters formed by the nodes.

In contrast to the networks with high similarity scores, the networks with low similarity scores in both human-written and ChatGPT-generated introduction sections (**Figure 6 (c) and (d)**) show more diverse dense nodes. The human-written introduction sections are centered on *traffic*, with other nodes such as *crashes, road, severity, factors,* and *vehicle*, and have clusters of *prediction, models, method, paper, used, modelling*, and *intersections, urban, ar-*

eas. In comparison, the ChatGPT-generated introduction sections have clusters of *traffic, factors road, accidents, severity, crashes*, and *safety, aim*, as well as a second cluster of *development, interventions, inform, policies, effective, targeted, reduce*, and *transportation*. This suggests that when given initial prompts, ChatGPT may generate more diverse clusters of nodes in its introduction sections compared to human-written introduction sections.

The results of the text networks in **Figure 6** suggest that the quality and diversity of introduction sections generated by ChatGPT may vary slightly depending on the strength of the prompts given. Additionally, the clusters of nodes formed by the introduction sections with initial prompts given to ChatGPT may differ from those generated by human-written introduction sections, indicating differences in interpretation and information processing between human writers and the ChatGPT model.

Table 4 presents the results of collocation frequency analysis of high and low similarity scores between the human-written and the initial prompt introduction section. The collocations with the highest frequency in the high similarity scores include *crash prediction models, road traffic crashes*, and *world health organization*, with counts of 13, 10, and 8, respectively. These collocations are also present in the initial prompt section, with *improve road safety, provide valuable insights,* and *interventions improve road* having the highest frequency counts of 193, 85, and 82, respectively.

Furthermore, the results show that the high similarity scores collocations contain technical terms and phrases relevant to road safety research, such as the *use seat belts, leading cause death*, and *motorcycle taxi riders*. These collocations are not present in the low similarity scores, mostly consisting of generic terms such as *road safety strategies, low-middle income countries*, and *crash severity prediction*. In addition, the counts of the collocations in the low similarity scores are relatively low compared to those in the high similarity scores, which may indicate a lack of relevant and specific information in the initial prompt.

The results of **Table 4** demonstrate that the high similarity scores contain technical and specific terms relevant to road safety research. In contrast, the low similarity scores have generic and less relevant terms. The high-frequency counts of the collocations for initial prompts in high and low similarity scores demonstrate how ChatGPT generates generic phrases in its outputs. On the other hand, the low frequency in the human written introduction and high similarity scores demonstrate different word choices among the various authors.

6 Conclusions and Study Limitations

This study evaluated the potential of ChatGPT to produce scholarly writeups that could be considered for publication. The introduction sections of 327 published articles were compared with the introductions prepared using ChatGPT. The ChatGPT-generated introductions were obtained using the papers' titles. With the application of the similarity analysis and text network analysis, the results were obtained and discussed. Based on the study findings, the following conclusions can be made.

- The analysis of the text networks for human-written and ChatGPT-generated research titles and introduction sections revealed key differences in the networks based on the strength of the prompts and the similarity scores.
- The human-written introduction sections tended to have more focused networks, with key nodes centered around the main topic of traffic. In contrast, the ChatGPT-generated networks showed more diverse and scattered nodes, indicating a broader range of top-

ics.

- Additionally, it is observed that the strength of the prompts had a relatively low impact on the resulting text networks. The improved prompts led to slightly more focused networks than initial prompts. The findings imply that the addition of a few pieces of information, such as a generic persona only, does not significantly affect the content of the output information. This finding suggests that the prompts' quality and specificity can be crucial in generating coherent and relevant text.
- Overall, the analysis highlights the importance of carefully considering the prompts used to generate text and the resulting text networks, as they can have a notable impact on the quality and coherence of the output. It is expected that the findings can contribute to the development of more effective and efficient text generation techniques, with a focus on producing high-quality, relevant, and coherent text.

Practically, this paper provides the basis for evaluating the capabilities of the prompts, which are the main components of the ChatGPT. The findings suggest that whenever a writer wants to retrieve information from ChatGPT, proper prompts with specific personas should be used. Failure to do so the ChatGPT outcomes will provide general information that is not significantly different from another person who utilized a nonspecific persona.

Despite the insights generated in this study, several gaps still need to be addressed to further understand the performance of ChatGPT's prompts in generating scholarly writeups.

- First, future studies may consider a set of specific personas in the respective domains. These personas can be well-known researchers in the respective domains. In this case, only the manuscripts from the respective people who are used as personas in the prompts are used. In addition to mentioning the specific persona, researchers can optimize the prompt by adding the specific details from the paper that they are interested in to evaluate whether ChatGPT would reproduce something similar to what that particular author wrote.
- Further, in addition to the metrics such as similarity scores, researchers may investigate the style of writing by asking technical writers to rate the ChatGPT-generated materials based on a set of criteria such as coherence, etc.
- Also, future studies may investigate the effectiveness of combining human and machine-generated text. This approach could lead to more coherent and relevant text while preserving the benefits of machine-generated text, such as efficiency and scalability. In conclusion, the study provides valuable insights into the impact of prompt strength on the coherence and relevance of text generated by ChatGPT. The findings call for further exploration of the gaps in this study and more research on the use of ChatGPT and other language models in scholarly writing. These studies could lead to more effective and efficient text-generation techniques and ultimately enhance the quality and rigor of scientific works.

References

Ali, M. J., & Djalilian, A. (2023). Readership Awareness Series – Paper 4: Chatbots and ChatGPT – Ethical considerations in scientific publications. *Seminars in Ophthalmology, 38* (5), 403–404. https://doi.org/10.1080/08820538.2023.2193444

Balat, A., & Bahsi, I. (2023). May artificial intelligence be a co–author on an academic paper?. *European Journal of Therapeutics*. https://api.semanticscholar.org/CorpusID:259573084

Bommarito, M. J., & Katz, D. M. (2022). GPT takes the bar exam. *SSRN Electronic Journal*. https://doi.org/

10.2139/SSRN.4314839

Chalkidis, I. (2023). ChatGPT may pass the bar exam soon, but has a long way to go for the LexGLUE benchmark. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4385460

Clarivate. (2023). *Web of Science Platform – Web of Science Group*. https://clarivate.com/webofsciencegroup/solutions/webofscience-platform/

Cooper, T. (2023). *Maximizing ChatGPT´s potential: The importance of quality prompts | Geeks*. https://vocal.media/geeks/maximizing-chat-gpt-s-potential-the-importance-of-quality-prompts

Dashti, M., Londono, J., Ghasemi, S., & Moghaddasi, N. (2023). How much can we rely on artificial intelligence chatbots such as the ChatGPT software program to assist with scientific writing?. *The Journal of Prosthetic Dentistry*. https://doi.org/10.1016/J.PROSDENT.2023.05.023

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2022). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *BioRxiv*, 2022.12.23.521610. https://doi.org/10.1101/2022.12.23.521610

Gao, C. A., Howard, F. M., Markov, N. S., Dyer, E. C., Ramesh, S., Luo, Y., & Pearson, A. T. (2023). Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. *NPJ Digital Medicine, 6* (1), 75. https://doi.org/10.1038/S41746-023-00819-6

Han, J., Kamber, M., & Pei, J. (2012). Getting to know your data. In J. Han, M. Kamber, & J. Pei (Eds.), *Data Mining* (pp. 39–82). Elsevier. https://doi.org/10.1016/B978-0-12-381479-1.00002-2

Hunter, S. (2014). A novel method of network text analysis. *Open Journal of Modern Linguistics*, *4*, 350–366. https://doi.org/10.4236/ojml.2014.42028

Jiang, C., Bhat, C. R., & Lam, W. H. K. (2020). A bibliometric overview of transportation research part B: Methodological in the past forty years (1979–2019). *Transportation Research Part B: Methodological, 138*, 268–291. https://doi.org/10.1016/j.trb.2020.05.016

Kocaballi, A. B. (2023). *Conversational AI-powered design: ChatGPT as designer, user, and product*. arXiv. https://doi.org/10.48550/arxiv.2302.07406

Kutela, B., Combs, T., John Mwekh'iga, R., & Langa, N. (2022). Insights into the long-term effects of COVID-19 responses on transportation facilities. *Transportation Research Part D: Transport and Environment, 111*, 103463. https://doi.org/10.1016/J.TRD.2022.103463

Kutela, B., Das, S., & Dadashova, B. (2021). Mining patterns of autonomous vehicle crashes involving vulnerable road users to understand the associated factors. *Accident Analysis & Prevention, 165*, 106473. https://doi.org/10.1016/J.AAP.2021.106473

Kutela, B., Dzinyela, R., Haule, H., Sheykhfard, A., & Msechu, K. (2023). Leveraging autonomous vehicles crash narratives to understand the patterns of parking-related crashes. *Traffic Safety Research, 4*, 000033. https://doi.org/10.55329/fiqq8731

Kutela, B., Kadeha, C., Magehema, R. T., Avelar, R. E., & Alluri, P. (2023). Leveraging text mining approach to explore research roadmap and future direction of wrong-way driving crash studies. *Data and Information Management, 2023*, 100044. https://doi.org/10.1016/J.DIM.2023.100044

Kutela, B., Kitali, A. E., & Kidando, E. (2022). Examining the influence of alternative fuels´ regulations and incentives on electric-vehicle acquisition. *International Conference on Transportation and Development*, 196–206. https://doi.org/10.1061/9780784484340.018

Kutela, B., Langa, N., Mwende, S., Kidando, E., Kitali, A. E., & Bansal, P. (2021). A text mining approach to elicit public perception of bike-sharing systems. *Travel Behaviour and Society, 24*, 113–123. https://doi.org/10.1016/j.tbs.2021.03.002

Kutela, B., Magehema, R. T., Langa, N., Steven, F., & Mwekh'iga, R. J. (2022). A comparative analysis of followers´ engagements on bilingual tweets using regression-text mining approach. A case of Tanzanian-based airlines. *International Journal of Information Management Data Insights, 2* (2), 100123. https://doi.org/10.1016/J.JJIMEI.2022.100123

Kutela, B., Msechu, K., Das, S., & Kidando, E. (2023). Chatgpt´s scientific writings: A case study on traffic safety. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4329120

Kutela, B., Msechu, K. J., Kidando, E., Das, S., & Kitali, A. E. (2023). Eliciting the influence of roadway and traffic conditions on hurricane evacuation decisions using regression–content analysis approach. *Travel Behaviour and Society, 33*, 100623. https://doi.org/10.1016/J.TBS.2023.100623

Kutela, B., Novat, N., Adanu, E. K., Kidando, E., & Langa, N. (2022). Analysis of residents´ stated preferences of shared micro–mobility devices using regression–text mining approach. *Transportation Planning and Technology, 45* (2), 159 – 178. https://doi.org/10.1080/03081060.2022.2089145

Kutela, B., Novat, N., & Langa, N. (2021). Exploring geographical distribution of transportation research themes related to COVID–19 using text network approach. *Sustainable Cities and Society, 67*, 102729. https://doi.org/10.1016/j.scs.2021.102729

Kutela, B., Oscar, C., Kidando, E., & Mihayo, M. (2023). Seeing through eyes of drivers: Space consideration in investigating visibility of Vulnerable road users involved in crashes from driver´s perspective. *Accident Analysis & Prevention, 192*, 107260. https://doi.org/10.1016/J.AAP.2023.107260

Kutela, B., & Teng, H. (2021). The use of Dynamic Message Signs (DMSs) on the freeways: An empirical analysis of DMSs logs and survey data. *Journal of Transportation Technologies, 11* (01), 90 – 107. https://doi.org/10.4236/jtts.2021.111006

Lahitani, A. R., Permanasari, A. E., & Setiawan, N. A. (2016). Cosine similarity to determine similarity measure: Study case in online essay assessment. *Proceedings of 2016 4th International Conference on Cyber and IT Service Management, CITSM 2016*. https://doi.org/10.1109/CITSM.2016.7577578

Liu, C., Bao, X., Zhang, H., Zhang, N., Hu, H., Zhang, X., & Yan, M. (2023). *Improving ChatGPT prompt for code generation*. https://arxiv.org/abs/2305.08360v1

Lu, C., Bu, Y., Wang, J., Ding, Y., Torvik, V., Schnaars, M., & Zhang, C. (2019). Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology, 70* (5), 462 – 475. https://doi.org/10.1002/ASI.24126

Macdonald, C., Adeloye, D., Sheikh, A., & Rudan, I. (2023). Can ChatGPT draft a research article? An example of population–level vaccine effectiveness analysis. *Journal of Global Health, 13*, 1–7. https://doi.org/10.7189/JOGH.13.01003

Mảjovsky, M., Černý, M., Kasal, M., Komarc, M., & Netuka, D. (2023). Artificial intelligence can generate fraudulent but authentic–looking scientific medical articles: Pandora´s box has been opened. *Journal of Medical Internet Research, 25*, e46924. https://doi.org/10.2196/46924

McCue, T. (2023). *ChatGPT Success completely depends on your prompt*. https://www.forbes.com/sites/tjmccue/2023/01/19/chatgpt–success–completely–depends–on–your–prompt/?sh=489a54f61a16

Newton, P. M., & Xiromeriti, M. (2023). ChatGPT performance on MCQ exams in higher education. *A pragmatic scoping review.* EdArXiv Preprints. https://doi.org/10.35542/OSF.IO/SYTU3

Noever, D., & McKee, F. (2023a). *Chatbots as problem solvers: Playing twenty questions with role reversals.* https://arxiv.org/abs/2301.01743v1

Noever, D., & McKee, F. (2023b). *Chatbots as Problem Solvers: Playing Twenty Questions with Role Reversals.* https://doi.org/10.48550/arxiv.2301.01743

OpenAI. (2020, June 11). *Generative pre–trained transformer*.

Paranyushkin, D. (2011). Identifying the pathways for meaning circulation using text network analysis. *Venture Fiction Practices, 2* (4). https://noduslabs.com/Research/Pathways–Meaning–Circulation–Text–Network–Analysis/

Perlman, A. (2022). The implications of ChatGPT for legal services and society. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.4294197

Probert, C. (2023). *Maximizing marketing efforts with ChatGPT: A guide to using ChatGPT for HubSpot content creation*. https://blog.whitehat–seo.co.uk/how–to–use–chat–gpt–with–hubspot

Schmid, S. (2023). *ChatGPT: How to write the perfect prompts – neuroflash.* https://neuroflash.com/chatgpt–how–to–write–the–perfect–prompts/

Selivanov, D. (2022). Text2vec: Modern text mining framework for R. *Comprehensive R Archive Network*

(CRAN). https://CRAN.R–project.org/package=text2vec

Sharma, G., & Thakur, A. (2023). *ChatGPT in drug discovery*. https://doi.org/10.26434/CHEM RXIV –2023 – QGS3K

Slater, D. (2023). *How to write better prompts for Chat GPT — GripRoom.* https://www.griproom.com/fun/how–to–write–better–prompts–for–chat–gpt

Zach. (2020). *How to calculate Jaccard Similarity in R*. https://www.statology.org/jaccard–similarity–in–r/

Zheng, H., & Zhan, H. (2023). ChatGPT in scientific writing: A cautionary tale. *The American Journal of Medicine, 136* (8), 725–726. https://doi.org/10.1016/J.AMJMED.2023.02.011

Zheng, O., Abdel–Aty, M., Wang, D., Wang, Z., & Ding, S. (2023). *ChatGPT is on the Horizon: Could a large language model be suitable for intelligent traffic safety research and applications?*. https://arxiv.org/abs/2303.05382v3