Sigma Journal of Engineering and Natural Sciences
Web page info: https://sigma.yildiz.edu.tr
DOI: 10.14744/sigma.2026.00000

Research Article

Using large language models for systematic literature reviews: A case study on electric vehicle routing problems

Emine Gülşen TORUNOĞLU ASLAN[1,*], Gonca YILDIRIM[2], Selçuk Kürşat İŞLEYEN[2], İsmail ASLAN[3]

[1]*Department of Supply Chain and Logistics, Graduate School of Natural and Applied Sciences, Gazi University, Ankara, 06560, Türkiye*
[2]*Department of Industrial Engineering, Faculty of Engineering, Gazi University, Ankara, 06560, Türkiye*
[3]*AltexSoft Inc., Foster City, California, 94404, USA*

ARTICLE INFO

Article history
Received: 12 August 2024
Revised: 02 November 2024
Accepted: 22 March 2025

Keywords:
Electric Vehicle Routing Problem; GPT-3.5; Large Language Models; Python LangChain Library; Systematic Literature Review; Text-Davinci-003

ABSTRACT

In academic research, systematic literature reviews play a key role in bringing together what is known about a given topic, but carrying out such reviews requires considerable time and effort, especially in fast-moving areas like logistics and transportation. Even the early task of locating and filtering relevant publications from the large body of available work can be slow and labor-intensive. Selection and interpretation of articles also remain prone to personal judgment, which threatens the reproducibility that scientific inquiry depends on. Against this backdrop, this study examines how advanced large language models can be incorporated into the systematic literature review process. To do so, we asked GPT-3.5 to screen research articles using a set of predefined criteria, testing it for both speed and accuracy. The input consisted of titles and abstracts from published work on the Electric Vehicle Routing Problem, a topic of growing importance in sustainable logistics, and each article was run through a carefully prepared prompt. GPT-3.5 scored 96% accuracy and finished the entire task in less than three minutes, whereas a PhD student working on the same task reached 84% accuracy over three working days. These outcomes suggest that large language models can be useful for automating parts or the whole of the literature review process. The workload can be reduced without sacrificing, and in fact even improving, accuracy. The study also provides a starting point for broader use of such models in academic research, illustrating how they may improve the speed and reproducibility of review work across different disciplines.

Cite this article as: Torunoğlu Aslan EG, Yıldırım G, İşleyen SK, Aslan İ. Using large language models for systematic literature reviews: A case study on electric vehicle routing problems. Sigma J Eng Nat Sci 2026;44(3):2233−2248.

INTRODUCTION

Artificial intelligence (AI) has found its way into many different fields, showing what it can do in areas ranging from project management and data analysis to various engineering tasks. For instance, AI has helped improve how workflows and decision-making processes are handled in

***Corresponding author.**
*E-mail address: egulsen.aslan@gazi.edu.tr

This paper was recommended for publication in revised form by Editor-in-Chief Ahmet Selim Dalkilic

Published by Yıldız Technical University Press, İstanbul, Turkey

construction project management [1], while methodologies like machine learning and blockchain have proven effective in log analysis [2]. More recently, a new wave of generative AI models has opened up fresh possibilities for how scientific research is carried out. Recent advancements in Natural Language Processing (NLP) have led to the development of large language models (LLM) that effectively generate human-like language and perform various language-related tasks. Although various research groups have studied LLMs for quite some time, only a narrow circle was aware of these advancements prior to the unveiling of ChatGPT on 30 November 2022.

Upon being made freely available on the Internet by OpenAI, a leading AI research laboratory, ChatGPT attracted immense interest, and the public suddenly became aware of the remarkable extent to which LLMs have advanced. ChatGPT is, in fact, a user interface to make people interact with Generative Pretrained Transformer (GPT) models at the back end, which are mainly built upon a deep neural network architecture known as transformers. GPT models were first introduced by Radford et al. [3] and are designed to generate human-like language by predicting the next word in a sequence of text based on the previous words [4].

The release of ChatGPT led to growing discussion about how LLMs might be used in areas ranging from education to business. Researchers also began exploring how these models support academic work [4, 5, 6], and literature reviews soon became one of the main areas of interest [4,5]. A literature review examines the existing research and scholarly publications related to a particular research topic or question. It is an essential component of academic and scientific research, which summarizes the current state of knowledge in the field and helps identify research gaps. However, due to the proliferation of scholarly articles, conducting such reviews has become increasingly demanding. Identifying relevant and high-quality sources among the large number of publications is time-consuming and challenging, and most of the time resource constraints impair the depth and comprehensiveness of the review. There is also a risk of personal bias and subjectivity in the selection and interpretation of studies, which may hinder reproducibility.

These challenges in the execution of comprehensive literature reviews have motivated researchers to develop automated methods to streamline and expedite the literature review process even before the LLMs. Notably, the tools, such as Fast2 [7, 8], Rayyan, SWIFT-Review and ASReview [9], were developed to support different stages of systematic literature review process. These tools use basic machine learning techniques, such as topic modeling and clustering, to assist in article screening and classification. Although these methods reduce manual effort and improve efficiency, they struggle with complex or domain-specific review tasks because they depend heavily on predefined algorithms and have limited contextual understanding. Scalability also becomes a challenge, especially when the datasets are large or contain highly specialized material.

LLMs have created new opportunities for addressing many of these shortcomings. Compared with earlier tools, LLMs provide a stronger ability to interpret contextual information and identify subtle relationships in text, allowing them to apply detailed inclusion and exclusion criteria consistently across large collections of articles. Prompt engineering plays an important role here, since it lets researchers modify selection criteria without retraining or fine-tuning the model. Despite these capabilities, the potential of LLMs in systematic literature reviews has not yet been sufficiently explored, especially in specialized or domain-specific settings.

The few studies that do exist have encountered several important limitations. Haman & Školník [10] used ChatGPT to conduct a literature review but came across numerous fabricated articles, a well-known problem called hallucination, in which the model produces simply wrong or nonexistent information. Rahman et al. [11] saw a similar pattern when they asked ChatGPT to compose an academic paper. The model generated fictitious citations and, when questioned later, acknowledged that it could not actually access the original articles. Both cases make clear that relying on LLMs like ChatGPT for literature reviews without structured guidance is problematic. To the best of the authors' knowledge, no peer-reviewed studies to date have systematically evaluated the application of advanced LLMs, such as GPT-3.5, in conducting systematic literature reviews, leaving a critical gap in understanding their potential and limitations in academic research workflows.

While the results of first studies may seem discouraging, they also highlight the need for more structured and guided methodologies when employing LLMs in systematic literature reviews. Having ChatGPT carry out a review on its own is only one approach to use an LLM. An alternative method involves working through the OpenAI API, which allows users to guide the model's behavior and make its responses more consistent with the task requirements. By providing structured prompts together with context-specific information, issues such as hallucination can be reduced, and the output can remain better aligned with the objectives of the study [12, 13]. This guided approach not only lowers the risk of incorrect outputs but also helps make systematic literature reviews more reproducible and reliable. The present study adopts this approach to evaluate the use of LLMs in the literature review process.

Building on this alternative approach, this study is, to the best of the authors' knowledge, the first to systematically evaluate the application of GPT-3.5, an advanced LLM, for systematic literature reviews. Unlike previous methods that relied on simpler machine learning tools or default configurations of LLMs, the methodology in this study employs structured input, prompt engineering, and domain-specific criteria to enhance accuracy and scalability. In doing so, it addresses important gaps in the existing literature and shows how LLMs can contribute to scientific workflows when applied with care.

To implement this approach, several techniques were used, including adding more context to the prompts, applying chain-of-thought reasoning, restricting the format of the outputs, and providing the model with supplementary data [12, 13]. In practice, the LLM receives the titles and abstracts of a predefined set of articles and evaluates their relevance to the research topic according to predetermined inclusion and exclusion criteria. Unlike earlier studies that relied mainly on the model's pretrained knowledge, the present study provides structured and task-specific input directly related to the review process. The OpenAI API is also used to limit the model to the supplied articles and stated criteria, which helps reduce hallucination and keeps the responses focused on the research task.

In this research, we test whether GPT-3.5 can successfully identify research publications related to the Electric Vehicle Routing Problem (EVRP). EVRP is a variant of the classical Vehicle Routing Problem (VRP), but includes more constraints due to extra features specific to electric vehicles. Electric vehicles have a restricted driving range, need extra time to recharge, and their energy consumption is affected by various factors. All these factors make solving EVRP more difficult than the classical VRP. Solving EVRP has also become more and more important due to growing environmental and regulatory demands. To find a reliable and effective solution requires first of all a comprehensive understanding of the existing literature.

Systematic literature reviews typically involve several structured steps to ensure a thorough analysis of existing literature. These include:

Step 1: Running a keyword search in relevant academic databases.

Step 2: Exporting the bibliographic details of articles listed in the search results.

Step 3: Screening the title and abstracts of the articles to determine their relevance based on predetermined inclusion/exclusion criteria.

Step 4: Investigating the full text of the selected articles to systematically classify them, summarizing existing knowledge and identifying research gaps.

The focus of this study is Step 3, namely the screening of titles and abstracts, which is a critical and time-intensive phase in systematic literature review studies. The research question is as follows:

Can an LLM effectively identify the articles that should be included in a systematic literature review by examining the titles and abstracts exported from an academic database?

The aim of this study is to evaluate whether GPT-3.5 can be used effectively in the systematic literature review process, particularly in the Electric Vehicle Routing Problem (EVRP). Accordingly, the study is guided by the following objectives:

- Assess the ability of GPT-3.5 to identify relevant articles using predefined inclusion and exclusion criteria.
- Compare the performance of GPT-3.5 with a human reviewer in terms of accuracy, precision, and recall.
- Evaluate whether GPT-3.5 can reduce the time required for article screening while maintaining effective performance.

MATERIALS AND METHODS

Figure 1 shows how GPT-3.5 is integrated into the systematic literature review process. To test can an LLM handle a systematic literature review, we collected a set of EVRP-related publications from the ProQuest academic database using the keyword string given below. The keyword string was kept deliberately broad so as not to miss any relevant EVRP papers.

noft(((("electric vehicle*" OR ev OR evs OR "green vehicle*" OR pollution OR "environmentally friendly vehicle*" OR "energy vehicle*" OR "recharging vehicle*" OR "alternative fuel vehicle*") AND routing AND problem*) OR (evr AND problem*) OR evrp OR "green VRP" OR gvrp OR "electric* travel salesman problem*") AND "electric* vehic*")*

The dataset exported from ProQuest comprises bibliographic information, including titles and abstracts of the articles published from 2011 until May 2023. After duplications were eliminated, 115 articles remained, all in English except for one in Turkish and one in Spanish. Three of the articles were missing abstracts. Manual efforts were undertaken to locate and complete the missing abstracts via other academic databases, ensuring that all 115 articles had both title and abstract information.

In the second step, upon quickly scanning through the titles and abstracts and considering the research objectives of the upcoming literature review study, inclusion and exclusion criteria listed in Table 1 were established to determine which articles should be selected as the studies focused on EVRP.

With the inclusion and exclusion criteria established, a PhD Student with prior experience studying EVRP was engaged to identify the relevant articles according to the inclusion and exclusion criteria by investigating their titles and abstracts in the extracted list obtained from ProQuest. The PhD Student was unaware that the results would be compared with any other result and made the selections with the understanding that the chosen articles would play a direct role in the forthcoming literature review study. The student was advised to complete the task immediately, but without setting an explicit deadline.

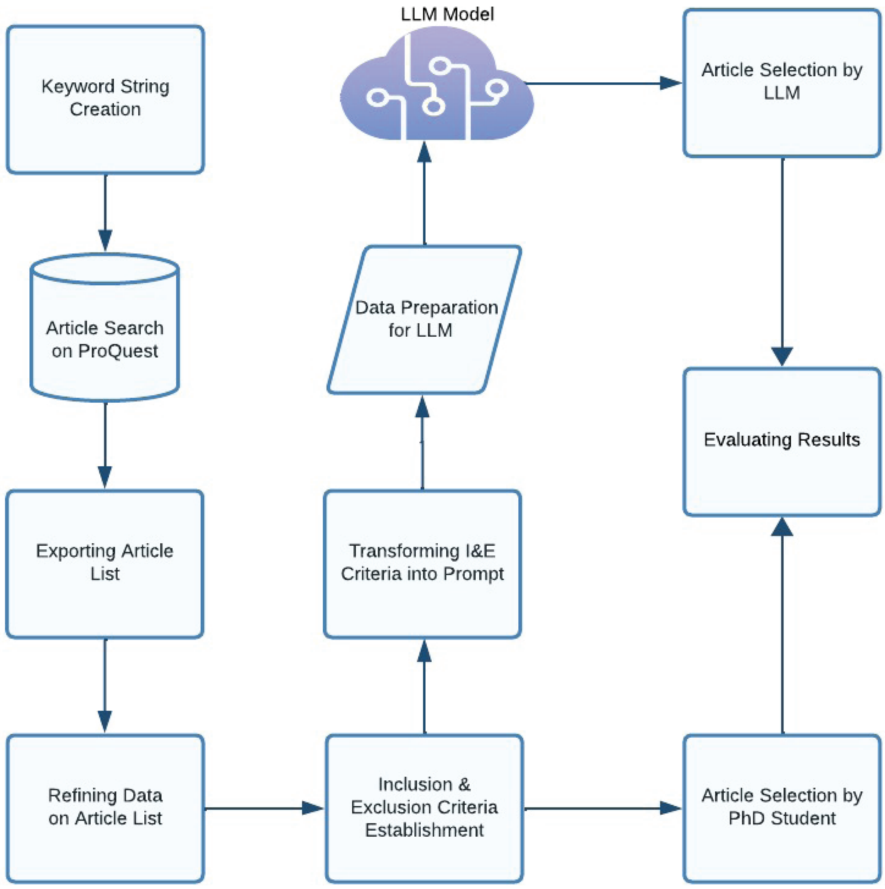

Figure 1. Methodology flowchart.

Table 1. Predetermined criteria to identify the articles

Inclusion Criteria	Exclusion Criteria
Electric bus and trucks	Powertrains
Electric vehicle transportation	Electric transportation
Logistics and transportation	Internet of Electric Vehicles (IoEV)
	Smart grid and Internet of Things (IoT) technologies
	Hybrid and combustion engine vehicles
	Charging infrastructure planning
	Ride-pooling and micro-transit
	Personal electric vehicles
	Single long trips
	Infrastructure systems and placement
	Recommendations to drivers and decision-making
	Equipment location
	Location routing
	Inventory routing
	Delivery routing
	Fixed route networks
	Diesel vehicles
	Transportation policy decision

Simultaneously, the inclusion and exclusion criteria were transformed into the prompt below to feed into the LLM:

query_string= "I am a researcher studying electric vehicle routing problem. I need to know if electric vehicle routing problem is studied in this article or not.

Details about my query are below:

- *I assume powertrains are not electric vehicles but electric bus and trucks are.*
- *I am interested in electric vehicle transportation but not in electric transportation.*
- *I am interested in logistics and transportation but not interested in*
  - *Internet of Electric Vehicles (IoEV),*
  - *smart grid and Internet of Things (IoT) technologies,*
  - *hybrid and combustion engine vehicles,*
  - *charging infrastructure planning,*
  - *ridepooling and microtransit,*
  - *personal electric vehicles,*
  - *single long trips,*
  - *infrastructure systems and placement,*
  - *recommendations to drivers, decision making,*
  - *equipment location, location routing,*
  - *inventory routing, delivery routing,*
  - *fixed-route networks, diesel vehicles,*
  - *and transportation policy decision.*

Give me an answer of yes or no."

Before feeding the dataset of articles and the prompt into the LLM, a data preparation phase was conducted. During this phase, the titles and abstracts of each paper were collected from the exported file, combined into a single cell, and saved as separate text files. Titles and abstracts in Turkish and Spanish were translated into English to ensure uniformity in the dataset.

Then, the Python LangChain Library [14] was employed to load each text file and to create a question-answering chain. The text-davinci-003 model, one of the LLMs under GPT-3.5 series, was utilized in the question-answering chain via OpenAI API. The text files fed into the model adhered to the constraint of its 4096-token limit, with each file either falling well below or precisely within this boundary. As shown in Table 2, all parameters were kept at their default settings except for the temperature, which was set to zero. This adjustment aimed to minimize the variability in the output and increase the consistency in the results, as recommended in the official OpenAI documentation [15].

Table 2. Model parameters

model_name='text-davinci-003'
temperature=0.0
max_tokens=256
top_p=1
frequency_penalty=0
presence_penalty=0
n=1
best_of=1

The question-answering chain takes the prompt and one of the saved text files, which has the title and abstract of one article, as inputs. Then, it yields an output of "yes" or "no" for the text file since the engineered prompt forces the model to provide a binary response.

RESULTS AND DISCUSSION

The PhD student completed the task over three working days, dedicating approximately five hours per day. On the other hand, the response time of the model was an average of 1.5 seconds with a standard deviation of 0.40 per article, so the LLM finished the task in less than three minutes. The articles selected by each are listed in Appendix A.

The number of articles selected by the PhD Student and LLM, as well as the ground truth values, are illustrated in the confusion matrices in Table 3 and Table 4 (see page 9). In the context of this research, the term "ground truth" refers to the authentic and validated data that serve as the baseline for evaluating the results of the student and LLM. The ground truth values are determined by two Subject Matter Experts and demonstrate the absolute true number of the articles that should be identified as topic related and not topic related.

Table 3. PhD student results

		PhD STUDENT		
		Topic Related	**Not Topic Related**	**TOTAL**
GROUND TRUTH	Topic Related	14	5	19
	Not Topic Related	13	83	96
	TOTAL	27	88	

Table 4. LLM results

		LLM		
		Topic Related	**Not Topic Related**	**TOTAL**
GROUND TRUTH	Topic Related	17	2	19
	Not Topic Related	3	93	96
	TOTAL	20	95	

A confusion matrix is a very useful tabular representation for binary classification problems, with two possible classes: "positive" and "negative" ("topic related" and "not topic related" in current case). It allows researchers to observe true positives, true negatives, false positives, and false negatives to calculate various performance metrics such as accuracy, precision, and recall [16]. In this study, a true positive (TP) indicates an instance that is actually "topic related" and is correctly classified as "topic related." Conversely, a true negative (TN) indicates an instance that is actually "not topic related" and is correctly classified as such. A false positive (FP) occurs when a "not topic related" instance is incorrectly classified as "topic related," whereas a false negative (FN) occurs when a "topic related" instance is incorrectly classified as "not topic related." As illustrated in Table 3, the confusion matrix for the PhD student yields TP = 14, FP = 13, TN = 83, and FN = 5, while Table 4 shows that the LLM's results are TP = 17, FP = 3, TN = 93, and FN = 2.

The performance metrics calculated for the PhD Student and LLM are illustrated in Table 5. Of these metrics, *accuracy* is the ratio of the total number of true positives and true negatives to the total number of articles (see Equation 1) and gives the overall correctness of the prediction. In Table 5, *accuracy* shows how many of the 115 articles are truly identified as topic related and not topic related by the PhD Student or LLM.

$$Accuracy = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}} \quad (1)$$

Recall is the ratio of true positives to the total number of true positives and false negatives (see Equation 2). It measures how well the prediction identifies the positive instances out of all the actual positive instances. In Table 5, *recall* shows how many of the topic related articles are truly identified as such by the PhD Student and LLM.

Table 5. Performance metrics

	PhD STUDENT	LLM
Accuracy	84%	96%
Recall	74%	89%
Precision	52%	85%

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \quad (2)$$

Finally, *precision* is the ratio of true positives to the total number of true positives and false positives (see Equation 3) and measures how many of the instances predicted as positive are true positives. It indicates how well the prediction avoids making false positive predictions. In Table 5, *precision* shows how many of the articles identified as topic related by the PhD Student and LLM are actually topic related.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \quad (3)$$

According to the metrics illustrated in Table 5, LLM has a strong potential to assist the literature review process, performing better at identifying the articles in the EVRP domain than the PhD Student. A comparison between the selections made by the text-davinci-003 model and those made manually by the Subject Matter Experts showed considerable overlap, suggesting that the model was generally successful in identifying relevant literature.

Therefore, the results of this study provide evidence that a*n LLM can effectively identify the articles that should be included in a systematic literature review by examining the titles and abstracts exported from an academic database.*

Screening titles and abstracts on similar topics is a tedious and demanding task, often requiring significant manual effort. Maintaining a consistent level of focus while evaluating the title and abstract of each article during a literature review study is significantly challenging for human researchers. Moreover, there is often a tendency to err on the side of including an article rather than excluding it when uncertainties arise, driven either by inadequately composed titles and abstracts or by lapses in concentration. An LLM can apply the same criteria consistently for each individual article, without the subjective biases inherent in human decision-making. Therefore, integrating LLMs into the article screening step helps reduce biases and support a more balanced and structured workflow.

The comparison between the PhD student and GPT-3.5 highlights the efficiency and quality of integrating LLMs into the literature review process. Both the PhD student and GPT-3.5 screened the same set of 115 articles based on predefined inclusion and exclusion criteria. The PhD student completed the task in approximately 24 hours over three working days, achieving an accuracy of 84%, recall of 74%, and precision of

52%. GPT-3.5 completed the same task in under three minutes with an accuracy of 96%, recall of 89%, and precision of 85%. These results show that GPT-3.5 reduces the time required for screening articles while at the same time achieving higher accuracy and better recall and precision.

The findings of this study demonstrate the capability of GPT-3.5 to substantially enhance the systematic literature review process in academic research. These results also highlight significant potential for engineering and industrial applications. In those applications where staying updated with the latest research is critical for innovation and problem-solving (e.g., transportation and logistics, renewable energy, robotics, manufacturing, materials science, etc.), engineers and industrial researchers can leverage LLMs in systematic literature reviews to streamline their research process and minimize manual effort. For instance, in the logistics and transportation sectors, the ability to rapidly identify relevant studies on electric vehicle routing can streamline fleet optimization strategies and support the adoption of sustainable transportation solutions. Similarly, in renewable energy systems engineering, this approach can accelerate the assessment of emerging technologies, enabling faster innovation and decision-making.

While GPT-3.5 demonstrated significant advantages in screening articles, the results of a traditional search were further examined by incorporating the inclusion and exclusion criteria (see Table 1) into the initial keyword string. For this purpose, a second keyword string shown below was prepared and a new search was conducted in the ProQuest database. This second query inexplicably yielded almost the same list of articles as the initial search. This result demonstrates that attempting a secondary query using Boolean operators to apply inclusion and exclusion criteria are not effective. These findings further underscore the value of LLMs in systematically and accurately applying inclusion and exclusion criteria during the article screening process.

noft((((“electric vehicle*” OR “electric* truck*” OR “electric* bus*” OR ev OR evs OR “green vehicle*” OR pollution OR “environmentally friendly vehicle*” OR “energy vehicle*” OR “recharging vehicle*” OR “alternative fuel vehicle*”) AND routing AND problem*) OR (evr AND problem*) OR evrp OR “green VRP” OR gvrp OR “electric* travel salesman problem*”) AND “electric* vehic*” NOT (powertrains AND “electric* transportation” AND “internet of electric vehicles” AND “smart grid” AND “Internet of Things” AND “hybrid vehicle*” AND “combustion engine” AND “charging infrastructure” AND ridepooling AND microtransit AND “personal electric vehicle*” AND “single trip” AND “infrastructure system*” AND “infrastructure placement” AND “recommendation* to driver*” AND “decision making” AND “equipment location” AND “location routing” AND “inventory routing” AND “delivery routing” AND “fixed-route network*” AND “diesel vehicle*” AND “transportation policy”))*

Challenges and Limitations

As the LLM for this study, GPT-3.5 was selected because it was affordable and widely accessible when the research was carried out. GPT-4 could potentially deliver better results thanks to its additional capabilities, but its associated costs were prohibitive for this research. It would be worth testing GPT-4 in future studies to see its potential to conduct systematic literature reviews.

On the methodological side, a zero-shot learning setup was adopted to evaluate the article screening capabilities of the LLM without additional training and validation procedure. The GPT-3.5 model, being a pretrained model, was used directly without additional fine-tuning or domain-specific training. Only prompt engineering was used to steer the behavior of the model and to keep its outputs relevant and consistent. Few-shot learning or fine-tuning on domain-specific data could, in principle, increase accuracy by giving the model tailored training or additional example; however, the main goal of this study was to assess what a pretrained LLM can do on its own. Whether few-shot learning or fine-tuning can improve the performance is a question for follow-up research.

Although this study encountered few practical challenges, there are still potential limitations to using LLMs in systematic literature reviews. One limitation is the reliance on prompt engineering to guide the model’s behavior. While no major problems arose in this study, working with less structured datasets or in unfamiliar domains could lead to greater variability in the outputs. The scalability of LLM-based methods also remains an open question, since the token limits constrain the volume of the text processed in a single query. It is also worth noting that this study dealt mainly with English-language abstracts. Applying the same approach to multilingual collections could introduce translation-related biases that would need careful attention in future work.

Due to time and resource constraints, this study was conducted with a dataset of 115 articles from a single academic database. Expanding the dataset to include articles from additional databases such as SCOPUS and Web of Science could make the findings more robust and reliable. On the human side, only one student carried out manual evaluation. While this still produced valuable insights, it does limit the generalization of the findings. Having several reviewers perform the same task in future studies would allow for a fairer comparison and stronger evidence.

Finally, the expertise of the human reviewer greatly influences performance. As a reviewer’s domain knowledge increases, human-level performance most probably improves, which could narrow the gap in accuracy, recall, and precision metrics. However, even in such a scenario, the LLM completes the task in less time, highlighting its efficiency and potential to expedite the systematic literature review process.

CONCLUSION

To the best of the authors' knowledge, this study is the first to integrate GPT-3.5 into systematic literature reviews, using tailored input and context-specific data to enhance outcomes, and to examine how large language models can be incorporated into academic review workflows. Using the Electric Vehicle Routing Problem as a representative case, this research investigates the use of large language models to screen article titles and abstracts for identifying relevant articles, thereby enhancing the efficiency and accuracy of the systematic literature review process.

The findings show that GPT-3.5 outperformed the human reviewer in terms of accuracy (96% vs. 84%), recall (89% vs. 74%), and precision (85% vs. 52%) while completing the task in under three minutes compared to three working days. According to these results, large language models can reduce the time and the manual effort needed for systematic literature reviews, without reducing the quality of the screening process, and in some cases improving it. GPT-3.5 performed noticeably better when it was given relevant data alongside carefully crafted prompts, which points to the importance of matching the model to the specific context of the task. This suggests that adapting the model to the context of the task can substantially improve performance compared with the default setup. Future work along these lines would help clarify just how far large language models can go in supporting literature reviews.

The success of this approach suggests that large language models can serve as a useful alternative to conventional manual methods for literature review workflows. The dataset used here was relatively small, yet the findings remain meaningful and clearly illustrate what large language models can do. They also lay a solid foundation for larger-scale studies and for branching out into fields beyond the Electric Vehicle Routing Problem. Researchers in healthcare, environmental science, logistics, and other areas where the volume of published work is expanding rapidly may benefit from a similar methodology. Moreover, because large language models apply inclusion and exclusion criteria the same way every time, they help reduce bias and strengthen reproducibility, both of which are central to sound scientific work.

Building on the results obtained with GPT-3.5, this study also provides a basis for future research on integrating large language models more extensively into systematic review workflows, including the direct querying of academic databases. Further progress in this area could make tasks like data retrieval and initial screening easier to handle and open the door to automating more of the process, leading to faster, more scalable, and more reproducible literature reviews.

NOMENCLATURE

AI	Artificial Intelligence
EVRP	Electric Vehicle Routing Problem
GPT	Generative Pretrained Transformer
LLM	Large Language Models
NLP	Natural Language Processing

AUTHORSHIP CONTRIBUTIONS

Authors equally contributed to this work.

DATA AVAILABILITY STATEMENT

The authors confirm that the data that supports the findings of this study are available within the article. Raw data that support the finding of this study are available from the corresponding author, upon reasonable request.

CONFLICT OF INTEREST

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ETHICS

There are no ethical issues with the publication of this manuscript.

STATEMENT ON THE USE OF ARTIFICIAL INTELLIGENCE

Artificial intelligence was not used in the preparation of the article.

REFERENCES

[1] Aladağ H, Güven İ, Balli O. Contribution of artificial intelligence (AI) to construction project management processes: State of the art with scoping review method. Sigma J Eng Nat Sci 2024;42:1654–1669. [CrossRef]

[2] Rahman RU, Kumar P, Kachare GP, Gawde MM, Tsundue T, Tomar DS. A novel machine learning-based artificial intelligence approach for log analysis using blockchain technology. Sigma J Eng Natural Sci 2024;42:1391–1409. [CrossRef]

[3] Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. OpenAI Tech Rep 2018;2018:1–12.

[4] Currie GM. Academic integrity and artificial intelligence: Is ChatGPT hype, hero or heresy? Seminars in Nucl Med 2023;53:719–730. [CrossRef]

[5] Alqahtani T, Badreldin HA, Alrashed M, Alshaya AI, Alghamdi SS, bin Saleh K, et al. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in Social and Administrative Pharmacy 2023;19:1236–1242. [CrossRef]

[6] Eke DO. ChatGPT and the rise of generative AI: Threat to academic integrity? J Respons Technol 2023;13:100060. [CrossRef]
[7] Yu Z, Carver JC, Rothermel G, Menzies T. Assessing expert system-assisted literature reviews with a case study. Exp Syst Appl 2022;200:116958. [CrossRef]
[8] Yu Z, Menzies T. FAST2: An intelligent assistant for finding relevant papers. Exp Syst Appl 2019;120:57–71. [CrossRef]
[9] Robledo S, Grisales Aguirre AM, Hughes M, Eggers F. "Hasta la vista, baby"-Will machine learning terminate human literature reviews in entrepreneurship? J Small Business Manag 2023;61:1314–1343. [CrossRef]
[10] Haman M, Školník M. Using ChatGPT to conduct a literature review. Accountabil Res 2023;30:1–3.
[11] Rahman M, Terano HJR, Rahman N, Salamzadeh A, Rahaman S. ChatGPT and academic research: A review and recommendations based on practical examples. J Educ Manag Dev Stud 2023;3:1–12. [CrossRef]
[12] Huyen C. Open challenges in LLM research. Chip Huyen Blog 2023;2023:1–10.
[13] Yan E. Patterns for building LLM-based systems & products. Eugene Yan Blog 2023;2023:1–15.
[14] Chase H. LangChain: Building applications with LLMs through composability. GitHub Repository 2022;2022:1–5.
[15] OpenAI. OpenAI API documentation and developer platform. OpenAI Tech Doc 2023;2023:1–10.
[16] James G, Witten D, Hastie T, Tibshirani R. Classification. In: An introduction to statistical learning: With applications in R. New York: Springer; 2013. p. 127–173. [CrossRef]
[17] Abousleiman RD. Energy-efficient routing for electric vehicles using metaheuristic optimization methods [doctoral dissertation]. Rochester: Oakland University; 2014. [CrossRef]
[18] Adler JD. Routing and scheduling of electric and alternative-fuel vehicles [doctoral dissertation]. Tempe: Arizona State University; 2014.
[19] Agárdi A, Kovács L, Bányai T. Ontology support for vehicle routing problem. Appl Sci 2022;12:12299. [CrossRef]
[20] Ahkamiraad A. Applications of optimization in vehicle routing problem and smart homes powered by local resources [doctoral dissertation]. Binghamton: New York; 2019.
[21] Alpos T, Iliopoulou C, Kepaptsoglou K. Nature-inspired optimal route network design for shared autonomous vehicles. Vehicles 2023;5:15–32. [CrossRef]
[22] Amini MH, Karabasoglu O. Optimal operation of interdependent power systems and electrified transportation networks. Energies 2018;11:196. [CrossRef]
[23] Ammous M. Optimal routing and charging of electric mobility-on-demand systems [master's thesis]. Moscow: University of Idaho; 2018. [CrossRef]
[24] Arani M, Rezvani MM, Davarikia H, Chan Y. Routing of electric vehicles in a stochastic network with non-recurrent incidents. arXiv preprint 2020;2020:1–12.
[25] Arnau Q, Juan AA, Serra I. On the use of learnheuristics in vehicle routing optimization problems with dynamic inputs. Algorithms 2018;11:208. [CrossRef]
[26] Awasthi A. Sustainable city logistics planning: Methods and applications. Hauppauge: Nova Science Publishers; 2019.
[27] Barco J, Guerra A, Muñoz L, Quijano N. Optimal routing and scheduling of charge for electric vehicles: A case study. Math Problems Eng 2017;2017:8509783. [CrossRef]
[28] Basso R. Route planning and energy consumption estimation for electric commercial vehicles [licentiate thesis]. Gothenburg: Chalmers University of Technology; 2017.
[29] Blum C, Liu H, Xiong H. CoordiQ: Coordinated Q-learning for electric vehicle charging recommendation. arXiv preprint 2021;2021:1–10.
[30] Bogyrbayeva A. Optimization and machine learning methods for solving combinatorial problems in urban transportation [doctoral dissertation]. Tampa: University of South Florida; 2021.
[31] Booth KEC. Constraint programming approaches to electric vehicle and robot routing problems [doctoral dissertation]. Toronto: University of Toronto; 2021.
[32] Bragin MA, Ye Z, Yu N. Toward efficient transportation electrification of heavy-duty trucks: Joint scheduling of truck routing and charging. arXiv preprint 2023;2023:1–12. [CrossRef]
[33] Bruchon MB. Analyzing and optimizing shared mobility fleet impacts [doctoral dissertation]. Pittsburgh: Carnegie Mellon University; 2022.
[34] Bruglieri M, Pezzella F, Pisacane O, Suraci S. A matheuristic for the electric vehicle routing problem with time windows. arXiv preprint 2015;2015:1–14. [CrossRef]
[35] Burkacky O, Mohr N, Pautasso L. Will quantum computing drive the automotive future? McKinsey Tech Rep 2020;2020:1–8.
[36] Caban J, Nieoczym A, Dudziak A, Krajka T, Stopková M. The planning process of transport tasks for autonomous vans-Case study. Appl Sci 2022;12:2993. [CrossRef]
[37] Çimen M, Belbağ S. Elektrikli araç varsayımı altında sürdürülebilir dinamik araç rotalama problemleri: Karma tamsayılı doğrusal programlama tabanlı bir çözüm algoritması. Third Sector Social Economic Review 2022;57:1572–1592. [CrossRef]
[38] Cubides LC, Arias Londoño A, Granada Echeverri M. Problema de ruteo de vehículos eléctricos con recogidas considerando la ubicación de estaciones de recarga y la operación del sistema de distribución de energía. TecnoLógicas 2019;22:15–32.

[39] da Costa LCA. Improvements on column-generation-based algorithms for vehicle routing and other combinatorial problems [doctoral dissertation]. Montreal: Polytechnique Montréal; 2020.

[40] Davatgari A, Cokyasar T, Subramanyam A, Larson J, Mohammadian A. Electric vehicle supply equipment location and capacity allocation for fixed-route networks. arXiv preprint 2023;2023:1–10.

[41] Dedeoglu M. Distributed learning and adaptive algorithms for edge networks [doctoral dissertation]. Tempe: Arizona State University; 2021.

[42] Ding N. Enabling urban parcel pickup and delivery services using all-electric trucks [doctoral dissertation]. Buffalo: State University of New York at Buffalo; 2017.

[43] Fang X. Analysis of electromagnetic interference problems caused by high-speed multiplayer boards with split reference plane [master's thesis]. Rolla: Missouri University of Science and Technology; 2022. [CrossRef]

[44] Fazeli SS. Stochastic programming models for electric vehicles' operation: Network design and routing strategies [doctoral dissertation]. Detroit: Wayne State University; 2020.

[45] Fazeli SS, Venkatachalam S, Smereka JM. Efficient algorithms for electric vehicles' min-max routing problem. Sustainable Operations and Computers 2023;4:110–118. [CrossRef]

[46] Fernández Gil A, Lalla-Ruiz E, Gómez Sánchez M, Castro C. A review of heuristics and hybrid methods for green vehicle routing problems considering emissions. Journal of Advanced Transportation 2022;2022:5714991. [CrossRef]

[47] Filipovska M. Travel time reliability in stochastic dynamic transportation networks: Modeling, path finding and routing [doctoral dissertation]. Evanston: Northwestern University; 2021.

[48] Furneaux A. Green logistics: Advanced methods for transport logistics management systems including platooning and alternative fuel powered vehicles [doctoral dissertation]. Canterbury: University of Kent; 2021.

[49] Gebbran D, Rich J, Dragicevic T. Electrical vehicle fleet routing accounting for dynamic battery degradation. arXiv preprint 2022;2022:1–10.

[50] Goodrich MT, Pszona P. Two-phase bicriterion search for finding fast and efficient electric vehicle routes. In: Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM; 2014. p. 193–202. [CrossRef]

[51] Chen H, Hatzenbühler J, Jenelius E. Pick-up and delivery problem for sequentially consolidated urban transportation with mixed and multi-purpose vehicle fleet. Journal of Advanced Transportation 2022;2022:2920532. [CrossRef]

[52] Zhang H. A WSN clustering multi-hop routing protocol using cellular virtual grid in IoT environment. Mathematical Problems in Engineering 2020;2020:8886687. [CrossRef]

[53] Haider Z. Using optimization methods for solving problems in sustainable urban mobility and conservation planning [doctoral dissertation]. Tampa: University of South Florida; 2020.

[54] Hong W, Zhang C, Chan C, Wang B. Electric vehicle charging infrastructure planning: A scalable computational framework. arXiv preprint 2020;2020:1–12.

[55] Houshmand A. Eco-routing and scheduling of connected and autonomous vehicles [doctoral dissertation]. Boston: Boston University; 2020.

[56] Huang JK, Tan Y, Lee D, Desaraju VR, Grizzle JK. Informable multi-objective and multi-directional RRT* system for robot path planning. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE; 2023. p. 5666–5673. [CrossRef]

[57] Hussain MA. V2G optimization in smart grid network [doctoral dissertation]. New York: The City College of New York; 2018.

[58] Gao J, Zheng X, Gao F, Tong X, Han Q. Heterogeneous multitype fleet green vehicle path planning of automated guided vehicle with time windows in flexible manufacturing system. Machines 2022;10:197. [CrossRef]

[59] Liu J. Design problems in distributed control of multi-agent systems [doctoral dissertation]. Notre Dame: University of Notre Dame; 2015.

[60] Zhang J, Yu M, Feng Q, Leng L, Zhao Y. Data-driven robust optimization for solving the heterogeneous vehicle routing problem with customer demand uncertainty. Complexity 2021;2021:6634132. [CrossRef]

[61] Jeong J, Ghaddar B, Zufferey N, Nathwani J. Adaptive robust electric vehicle routing under energy consumption uncertainty. arXiv preprint 2022;2022:1–12.

[62] Jiang Y, Yang R, Zang C, Wei Z, Thompson J, Tran TH, et al. Toward baggage-free airport terminals: A case study of London City Airport. Sustainability 2022;14:212. [CrossRef]

[63] Ju B, Kim M, Moon I. Vehicle routing problem considering reconnaissance and transportation. Sustainability 2021;13:3188. [CrossRef]

[64] Jung J. Shared-ride passenger transportation systems with real-time routing [doctoral dissertation]. Los Angeles: University of Southern California; 2012.

[65] Karamete BK, Glaser E. Optimal routing algorithm for trips involving thousands of EV-charging stations using Kinetica-Graph. arXiv preprint 2022;2022:1–10.

[66] Khayati Y. Household use of emerging vehicle technologies: Modeling framework and traveler adaptation [doctoral dissertation]. Buffalo: State University of New York at Buffalo; 2018.

[67] Kim G. Dynamic vehicle routing problem with fuzzy customer response. Sustainability 2023;15:4376. [CrossRef]

[68] Kosmanos D, Maglaras LA, Mavrovouniotis M, Moschoyiannis S, Argyriou A, Maglaras A, et al. Route optimization of electric vehicles based on dynamic wireless charging. IEEE Access 2018;6:42551–42565. [CrossRef]

[69] Chen L, Ma H, Wang Y, Li F. Vehicle routing problem for the simultaneous pickup and delivery of lithium batteries of small power vehicles under charging and swapping mode. Sustainability 2022;14:9883. [CrossRef]

[70] Li L. Optimal mixed bus fleet management strategy for emissions reduction [doctoral dissertation]. Hong Kong: Hong Kong University of Science and Technology; 2018.

[71] Wang L, Ma W, Wang L, Ren Y, Yu C. Enabling in-depot automated routing and recharging scheduling for automated electric bus transit systems. J Adv Trans 2021;2021:5531063. [CrossRef]

[72] Lam AYS, Li VOK. Opportunistic routing for vehicular energy network. IEEE Int Things J 2018;5:533–545. [CrossRef]

[73] Lian Y, Lucas F, Sörensen K. The electric on-demand bus routing problem with partial charging and nonlinear function. SSRN Scholarly Paper 2022;2022:1–25. [CrossRef]

[74] Liang Y, Dabia S, Luo Z. The electric vehicle routing problem with nonlinear charging functions. arXiv preprint 2021;2021:1–15.

[75] Rahman MA, Shahriar MH, Al-Shaer E, Zhu Q. A formal approach for efficient navigation management of hybrid electric vehicles on long trips. arXiv preprint 2019;2019:1–12.

[76] Alqahtani M, Hu M. Integrated energy scheduling and routing for a network of mobile prosumers. Energy 2020;200:117451. [CrossRef]

[77] Ma D. A geometric optimization approach to the location and routing problems in modern economy [doctoral dissertation]. Boston: Northeastern University; 2021.

[78] Maghfiroh MFN, Yu VF, Redi AANP, Abdallah BN. A location routing problem with time windows consideration: A metaheuristics approach. Appl Sci 2023;13:843. [CrossRef]

[79] Maglaras LA, Jiang J, Maglaras AL, Topalis FV. Mobile energy disseminators increase electrical vehicles range in a smart city. IET Conference Publications 2014;2014:3.3. [CrossRef]

[80] Matin Moghaddam N. The partially rechargeable electric vehicle routing problem with time windows and capacitated charging stations [master's thesis]. Clemson: Clemson University; 2015.

[81] Meenaakshi Sundhari RP, Murali L, Baskar S, Shakeel PM. MDRP: Message dissemination with re-route planning method for emergency vehicle information exchange. Network Appl 2021;14:2285–2294. [CrossRef]

[82] Min X. Modelling and optimization for battery electric vehicle operations [doctoral dissertation]. Singapore: National University of Singapore; 2018.

[83] Moradipari A. Learning and pricing algorithms for human-cyber-physical systems [doctoral dissertation]. Santa Barbara: University of California, Santa Barbara; 2022.

[84] Movahed Nejad M. Frontiers in operations research for overcoming barriers to vehicle electrification [doctoral dissertation]. Detroit: Wayne State University; 2015.

[85] Muthu KP. An investigation of inventory routing problem with carbon emission and sustainability consideration [master's thesis]. Buffalo: State University of New York at Buffalo; 2021.

[86] Narayanan A, Misra P, Ojha A, Bandhu V, Ghosh S, Vasan A. A reinforcement learning approach for electric vehicle routing problem with vehicle-to-grid supply. arXiv preprint 2022;2022:1–10. [CrossRef]

[87] Ni L. Modeling and optimization of electric vehicle networks in future sustainable cities [doctoral dissertation]. Hong Kong: Hong Kong University of Science and Technology; 2021.

[88] Normasari NME, Yu VF, Bachtiyar C, Sukoyo U. A simulated annealing heuristic for the capacitated green vehicle routing problem. Math Problems Eng 2019;2019:2358258. [CrossRef]

[89] Nosrati M, Arshadi Khamseh A. Distance discount in the green vehicle routing problem offered by external carriers. SN Appl Sci 2020;2:1443. [CrossRef]

[90] Nyotta RBS. Essays on revenue management in urban mobility [doctoral dissertation]. Los Angeles: University of California, Los Angeles; 2021.

[91] Oliveira LK, Sousa LTM, Nascimento COL, Silva TGC, Pinto PHG, Garcia MN, et al. Sustainable city logistics planning: Methods and applications. Vol. 2. Hauppauge: Nova Science Publishers; 2020.

[92] Zhou P. Building high performance wireless rechargeable sensor networks and vehicular networks by algorithm designs [doctoral dissertation]. Stony Brook: State University of New York at Stony Brook; 2020.

[93] Paparella F, Chauhan K, Hofman T, Salazar M. Electric autonomous mobility-on-demand: Joint optimization of routing and charging infrastructure siting. arXiv preprint 2022;2022:1–12.

[94] Li Q. Optimal routing problems in stochastic transportation networks considering sustainability [doctoral dissertation]. Evanston: Northwestern University; 2015.

[95] Qiu J, Du L. A charging-as-a-service platform for charging electric vehicles on the move: New vehicle routing model and solution. arXiv preprint 2021;2021:1–14.

[96] Quilliot A, Sarbinowski A, Toussaint H. Vehicle driven approaches for non-preemptive vehicle relocation with integrated quality criterion in a vehicle sharing system. Ann Oper Res 2021;298:445–468. [CrossRef]
[97] Quintero E, Sanchez M, Roldan N, Toro M. Genetic algorithm for optimal distribution in cities. arXiv preprint 2018;2018:1–8.
[98] Raeesi R. Mathematical models and solution algorithms for the vehicle routing problem with environmental considerations [doctoral dissertation]. Lancaster: Lancaster University; 2019.
[99] Rajan P. Electric vehicle route planning in the presence of stochasticity [doctoral dissertation]. Riverside: University of California, Riverside; 2022.
[100] Rajesh K, Jain E, Kotecha P. A multi-objective approach to the electric vehicle routing problem. arXiv preprint 2022;2022:1–10.
[101] Rodríguez-Esparza E, Masegosa AD, Oliva D, Onieva E. A new hyper-heuristic based on adaptive simulated annealing and reinforcement learning for the capacitated electric vehicle routing problem. arXiv preprint 2022;2022:1–12.
[102] Roselli SF. Conflict-free routing of mobile robots [doctoral dissertation]. Gothenburg: Chalmers University of Technology; 2022.
[103] Wang S. A metaheuristic for the electric vehicle routing problem with recharging stations and time windows [master's thesis]. Clemson: Clemson University; 2015.
[104] Schneider M, Stenger A, Goeke D. The electric vehicle-routing problem with time windows and recharging stations. Transport Sci 2014;48:500–520. [CrossRef]
[105] Sciomachen A, Truvolo M. An exact approach for selecting pickup-delivery stations in urban areas to reduce distribution emission costs. Mathematics 2023;11:1876. [CrossRef]
[106] Shao S. Market design for next generation of shared and electric transportation systems: Modeling, optimization, and learning [doctoral dissertation]. Columbus: The Ohio State University; 2022.
[107] Shi J. Intelligent control and data-driven algorithms for critical infrastructure systems [doctoral dissertation]. Riverside: University of California, Riverside; 2021.
[108] Socarrás Bertiz CA, Fernández Lozano JJ, Gomez-Ruiz JA, García-Cerezo A. Integration of a mobile node into a hybrid wireless sensor network for urban environments. Sensors 2019;19:215. [CrossRef]
[109] Sohet B, Hayel Y, Beaude O, Jeandin A. Hierarchical coupled routing-charging model of electric vehicles, stations and grid operators. IEEE Transactions on Smart Grid 2021;12:5146–5157. [CrossRef]
[110] Song Y. A novel location-allocation-routing model for siting multiple recharging points on the continuous network space [doctoral dissertation]. Tempe: Arizona State University; 2020.
[111] Swaminathan D, Rajagopalan A. Multi-objective golden flower optimization algorithm for sustainable reconfiguration of power distribution network with decentralized generation. Axioms 2023;12:70. [CrossRef]
[112] Sweda TM. Selected topics on decision making for electric vehicles [doctoral dissertation]. Evanston: Northwestern University; 2014.
[113] Wang T. Control and optimization approaches for power management in energy-aware battery-powered systems [doctoral dissertation]. Boston: Boston University; 2013.
[114] Tang X, Sun C, Bi S, Wang S, Zhang AY. A holistic review on advanced bi-directional EV charging control algorithms. ACM SIGEnergy Energy Inform Revi 2021;1:78–88. [CrossRef]
[115] Tookanlou MB. Optimal day-ahead scheduling frameworks for e-mobility ecosystem operation with drivers preferences under uncertainties [doctoral dissertation]. Newcastle upon Tyne: University of Northumbria; 2022.
[116] Tou WH. Metaheuristics for solving variants of vehicle routing problems [doctoral dissertation]. Hong Kong: The Chinese University of Hong Kong; 2016.
[117] Tu R. Traffic emission modeling for robust policy design in connected and electric transportation [doctoral dissertation]. Toronto: University of Toronto; 2020.
[118] Umeda Y. Special issue on design and manufacturing for environmental sustainability. Int J Autom Technol 2020;14:855–856. [CrossRef]
[119] Unno M, Mizuno S. Effective computation environment for the traveling salesman problem using three-dimensional information. Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA) 2017;2017:265–269.
[120] Yu VF, Nguyen MPK, Putra K, Gunawan A, Dharma IGBB. The two-echelon vehicle routing problem with transshipment nodes and occasional drivers: Formulation and adaptive large neighborhood search heuristic. J Adv Transport 2022;2022:5603956. [CrossRef]
[121] Zhou W. Green vehicle routing problems in the new mobility era [doctoral dissertation]. Chicago: University of Illinois at Chicago; 2016.
[122] Liu X. Computational methods for solving the combinatorial optimization problems in transportation [doctoral dissertation]. Tampa: University of South Florida; 2022.
[123] Wang YD, Lu XC, Song YM, Feng Y, Shen JR. Monte Carlo tree search improved genetic algorithm for unmanned vehicle routing problem with path flexibility. Adv Product Eng Manag 2022;17:425–438. [CrossRef]

[124] Li Y, Peyman M, Panadero J, Juan AA, Xhafa F. IoT analytics and agile optimization for solving dynamic team orienteering problems with mandatory visits. Mathematics 2022;10:982. [CrossRef]
[125] Yao C, Chen S, Salazar M, Yang Z. Joint routing and charging problem of electric vehicles with incentive-aware customers considering spatio-temporal charging prices. IEEE Trans Intell Transport Syst 2023;24:14512–14523. [CrossRef]
[126] Yi Z. Optimizing battery electric transportation: Energy and infrastructure considerations [doctoral dissertation]. Notre Dame: University of Notre Dame; 2017.
[127] Yuan Y, Li H, Ji L. Application of deep reinforcement learning algorithm in uncertain logistics transportation scheduling. Comput Intell Neurosci 2021;2021:5672227. [CrossRef]
[128] Gao Z, Ye C. Reverse logistics vehicle routing optimization problem based on multivehicle recycling. Math Problems Eng 2021;2021:5559684. [CrossRef]
[129] Li Z. Development and evaluation of an intelligent transportation systems-based architecture for electric vehicles [doctoral dissertation]. Clemson: Clemson University; 2013.
[130] Zalesak MDN. Models, algorithms, and implementations for operational optimization of ridepool services [doctoral dissertation]. Ithaca: Cornell University; 2021.
[131] Zhao Y, Zeng T, Allybokus Z, Guo Y, Moura S. Joint design for electric fleet operator and charging service provider: Understanding the non-cooperative nature. IEEE Transactions on Intelligent Transport Syst 2023;24:115–127. [CrossRef]

APPENDIX A. List of the Examined Studies

	PhD Student	LLM	Ground Truth
Abousleiman [17]	1	0	0
Adler [18]	1	0	0
Agárdi et al. [19]	0	0	0
Ahkamiraad [20]	0	0	0
Alpos et al. [21]	0	0	0
Amini & Karabasoglu [22]	0	0	0
Ammous [23]	0	0	0
Arani et al. [24]	1	0	0
Arnau et al. [25]	0	0	0
Awasthi, A. [26]	0	0	0
Barco et al. [27]	1	0	0
Basso [28]	1	0	1
Blum et al. [29]	0	0	0
Bogyrbayeva [30]	0	0	0
Booth [31]	0	0	0
Bragin et al. [32]	1	1	1
Bruchon [33]	0	0	0
Bruglieri et al. [34]	1	1	1
Burkacky et al. [35]	0	0	0
Caban et al. [36]	0	0	0
Çimen & Belbağ [37]	1	1	1
Cubides et al. [38]	0	1	1
da Costa [39]	0	0	0
Davatgari et al. [40]	0	0	0
Dedeoglu [41]	0	0	0
Ding [42]	0	1	1
Fang [43]	0	0	0
Fazeli [44]	1	0	0
Fazeli et al. [45]	1	1	1
Fernández Gil et al. [46]	0	0	0
Filipovska [47]	0	0	0
Furneaux [48]	0	0	0
Gebbran et al. [49]	1	1	1
Goodrich & Pszona [50]	0	1	1
H. Chen et al. [51]	0	0	0
H. Zhang [52]	0	0	0
Haider [53]	0	0	0
Hong et al. [54]	0	0	0
Houshmand [55]	0	0	0
Huang et al. [56]	0	0	0
Hussain [57]	0	0	0
J. Gao et al. [58]	0	0	0
J. Liu [59]	0	0	0
J. Zhang et al. [60]	0	0	0
Jeong et al. [61]	1	0	1
Jiang et al. [62]	0	0	0

APPENDIX A. List of the Examined Studies

	PhD Student	LLM	Ground Truth
Ju et al. [63]	0	0	0
Jung [64]	0	0	0
Karamete & Glaser [65]	1	0	0
Khayati [66]	0	0	0
Kim [67]	1	0	0
Kosmanos et al. [68]	1	0	0
L. Chen et al. [69]	0	0	0
L. Li [70]	0	0	0
L. Wang et al. [71]	0	0	0
Lam & Li [72]	0	0	0
Lian et al. [73]	1	0	0
Liang et al. [74]	1	1	1
M. A. Rahman et al. [75]	0	0	0
M. Alqahtani & Hu [76]	0	0	0
Ma [77]	0	0	0
Maghfiroh et al. [78]	0	0	0
Maglaras et al. [79]	0	0	0
Matin Moghaddam [80]	1	1	1
Meenaakshi Sundhari et al. [81]	0	0	0
Min [82]	0	0	0
Moradipari [83]	0	0	0
Movahed Nejad [84]	0	0	0
Muthu [85]	0	0	0
Narayanan et al. [86]	1	1	1
Ni [87]	0	0	0
Normasari et al. [88]	0	0	0
Nosrati & Arshadi Khamseh [89]	0	0	0
Nyotta [90]	0	0	0
Oliveira et al. [91]	0	0	0
P. Zhou [92]	0	0	0
Paparella et al. [93]	0	0	0
Q. Li [94]	0	0	0
Qiu & Du [95]	1	1	0
Quilliot et al. [96]	0	0	0
Quintero et al. [97]	0	1	1
Raeesi [98]	0	0	0
Rajan [99]	1	0	0
Rajesh et al. [100]	0	1	0
Rodríguez-Esparza et al. [101]	1	1	1
Roselli [102]	1	0	0
S. Wang [103]	1	1	1
Schneider et al. [104]	1	1	1
Sciomachen & Truvolo [105]	0	0	0
Shao [106]	0	0	0
Shi [107]	0	0	0
Socarrás Bertiz et al. [108]	0	0	0

APPENDIX A. List of the Examined Studies

	PhD Student	LLM	Ground Truth
Sohet et al. [109]	0	0	0
Song [110]	0	0	0
Swaminathan & Rajagopalan [111]	0	0	0
Sweda [112]	0	0	0
T. Wang [113]	0	0	0
Tang et al. [114]	0	0	0
Tookanlou [115]	0	1	1
Tou [116]	0	0	0
Tu [117]	0	0	0
Umeda [118]	0	0	0
Unno & Mizuno [119]	0	0	0
V. F. Yu et al. [120]	1	0	0
W. Zhou [121]	1	1	1
X. Liu [122]	0	0	0
Y. D. Wang et al. [123]	0	0	0
Y. Li et al. [124]	0	0	0
Yao et al. [125]	0	0	0
Yi [126]	0	0	0
Yuan et al. [127]	0	0	0
Z. Gao & Ye [128]	0	0	0
Z. Li [129]	0	1	0
Zalesak [130]	0	0	0
Zhao et al. [131]	0	0	0