ICCEPM 2024 https://dx.doi.org/10.6106/ICCEPM.2024.0407
The 10th International Conference on Construction Engineering and Project Management
Jul. 29-Aug.1, 2024, Sapporo

Identification of Social Conflict Stakeholders in Public Infrastructure Projects using ChatGPT

Do Namgoong[1]*, Seung H. Han[2]

[1] *Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, Korea,*
E-mail address: 821david@yonsei.ac.kr
[2] *Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, Korea,*
E-mail address: shh6018@yonsei.ac.kr

Abstract: Social conflict surrounding public infrastructure projects has grown because of increasing project complexity and accelerated conflict propagation. This social conflict stems from various concerns ranging from environmental issues to regulatory and compliance requirements resulting in the intervention of various stakeholders with different interests. Against this backdrop, understanding the stakeholders involved and their dynamics is crucial for effective project management and smooth implementation of the project. Therefore, this paper introduces an analytical process utilizing ChatGPT to automatically identify stakeholders involved in the social conflict surrounding the public infrastructure project from news articles. As a result, a stakeholder network is constructed to delve into the complex stakeholder interrelationships and identify key stakeholders of the specific period. To explore the potential of the proposed process, an experimental case study of the Jeju 2nd Airport project, which suffered from a high level of social conflict, was conducted. The proposed process enables timely analysis of the conflict situation which is crucial for successful conflict management. This study highlights the significance of a systemic approach to timely stakeholder analysis, setting the groundwork for a quantitative and up-to-date investigation of social conflicts around public infrastructure projects.

Key words: conflict management, public infrastructure projects, ChatGPT, network analysis, entity extraction

1. INTRODUCTION

1.1. Research Background

In recent years, the environment surrounding public infrastructure projects has seen significant changes, characterized by heightened complexity and the rapid propagation of social conflict. Public infrastructure projects often encompass a broad spectrum of concerns, ranging from environmental impact to urban development, and should go through a labyrinth of regulatory and compliance requirements [1-2]. Therefore, these projects inherently affect various stakeholders including government entities, private sector participants, local communities, and environmental groups. Moreover, a proliferation of real-time communication technologies accelerated this phenomenon [3]. These technologies have democratized information dissemination, enabling a wider array of individuals and groups to engage in discourse about public infrastructure projects. As a result, they catalyzed more frequent and intensified social conflicts. This phenomenon poses significant challenges for project managers and policymakers, as more frequent and intensified social conflicts can substantially obstruct project implementation [4]. Consequently, there arises a necessity to establish a process to analyze and manage these social conflicts effectively.

In this context, news platforms play a pivotal role in disseminating information about various social conflicts related to public infrastructure projects [2]. As a comprehensive repository of social conflicts, news articles provide valuable insights to understand the dynamics of public opinion and stakeholder

concerns. Extracting relevant information from vast amounts of textual data requires sophisticated analytical tools, and Natural Language Processing (NLP) technologies are widely used tools for this purpose. Among the emerging tools in this field, ChatGPT has gained attention for its ability to process and analyze text data efficiently without extra fine-tuning for specific tasks [5-6]. Using ChatGPT, the extraction of nuanced information from text sources like news is possible, offering an opportunity to understand and navigate the intricate network of social conflicts surrounding public infrastructure projects.

1.2. Research Objective

The primary objective of this research is to develop and validate a process for the automated analysis of stakeholder dynamics and interrelationships in the context of social conflicts surrounding public infrastructure projects. This process leverages the capabilities of ChatGPT to navigate the complexities of stakeholder analysis. The objectives of this study can be primarily categorized into two areas. The first objective is the systematic identification of stakeholders involved in conflicts surrounding public infrastructure projects using news data. For this purpose, the study utilizes ChatGPT to analyze data obtained from various news sources to deduce stakeholders related to the conflict. The second objective involves analyzing the interactions and complexities of relationships among the identified stakeholders. This analysis aims to determine which stakeholders play pivotal roles in the conflicts surrounding the projects. This research not only aims to contribute to the academic understanding of stakeholder dynamics in public infrastructure projects but also seeks to provide practical tools for project managers and policymakers to effectively navigate and manage these complex social landscapes surrounding public infrastructure projects. Additionally, we hope that this research will lay the foundation for studies that seek a profound understanding of conflict situations through the utilization of textual data such as news data.

2. LITERATURE REVIEW

As social conflicts become more prevalent, a growing body of research has emerged, focusing on various aspects of conflict management. Olander [7] pioneered an analytical framework for stakeholders in construction projects, focusing on their influence and likelihood of exerting it. Utilizing a stakeholder impact/probability matrix, this study developed the 'Stakeholder Impact Index,' a tool for evaluating stakeholder characteristics in conflict scenarios. In a similar vein, Elias [8] introduced a comprehensive system for stakeholder analysis in environmental planning and management. This approach entailed identifying each stakeholder's interests and applying system dynamics methodology to understand the dynamic nature of conflicts. Li et al. [9] proposed a multi-factor hierarchical fuzzy comprehensive evaluation model, emphasizing the crucial role of stakeholder satisfaction. They argued that neglecting stakeholder concerns could precipitate project failures, advocating for a fuzzy logic-based approach to incorporate stakeholder participation in evaluations. These studies collectively underscore the importance of intricate stakeholder analysis in conflict management within construction.

The complexity of stakeholder relationships has led researchers to embrace Social Network Analysis (SNA) as an insightful methodology. SNA offers a systematic framework for dissecting the intricate web of stakeholder interconnections. Lienert et al. [10] employed network analysis with a specific focus on the planning phase of water infrastructure projects, where social conflicts are prevalent. Their study aimed to discern stakeholder representation across various decision-making levels and sectors, utilizing semi-structured interviews for data collection. Mok et al. [11] applied network theory to explore the nexus of stakeholder-related issues in a large-scale building project. They tried to map out the network of stakeholder issues, identifying key stakeholders, and elucidating the interrelationships among them. Their analysis yielded crucial insights into pivotal issues and relationships that influence project execution. Xue et al. [12] advanced this line of inquiry by proposing a network-based framework for examining stakeholder conflict dynamics. Their work identified key conflicts and affected relationships and proposed a mirror-Z approach, which considered the conflict criticalness and stakeholder participation of conflicts, to suggest management priority on stakeholder conflicts for the practitioners. While past studies have focused on deriving key stakeholders and issues, this study is meaningful in that their study presented a comprehensive framework that enables the analysis and derivation of management studies for practitioners.

However, the reliance on interview-based data in SNA poses limitations, such as biases and the inability to analyze real-time situations. Moreover, a notable limitation of the existing conflict

management frameworks is their predominantly retrospective nature. This retrospective focus limits the ability of practitioners to proactively address issues as they emerge, emphasizing the need for more dynamic and real-time methodologies. There is a growing need for real-time, objective data sources, such as news articles, to monitor stakeholder dynamics and conflicts. However, such an approach remains underexplored in the literature, signaling a potential room for future research.

3. METHODOLOGY

This study aims to propose a framework for automatically analyzing stakeholders and their interrelationships regarding social conflicts using ChatGPT. This process encompasses four primary functions: web crawling, entity extraction, entity matching, and network analysis. Initially, web crawling is employed to collect Korean news articles related to the project under analysis. Subsequently, ChatGPT is utilized for the extraction of entities representing stakeholders involved in the project's conflict. Because these entities may be referred to in multiple forms across the collected news articles, an entity matching process is implemented to consolidate duplicate entities under varying expressions. After matching the entity from the entire dataset, a comprehensive stakeholder network is constructed based on the matched entities across the entire dataset. The entire process is presented in Figure 1 below. Detailed explanations of these methodologies, along with their backgrounds, are elaborated in the following subsections.

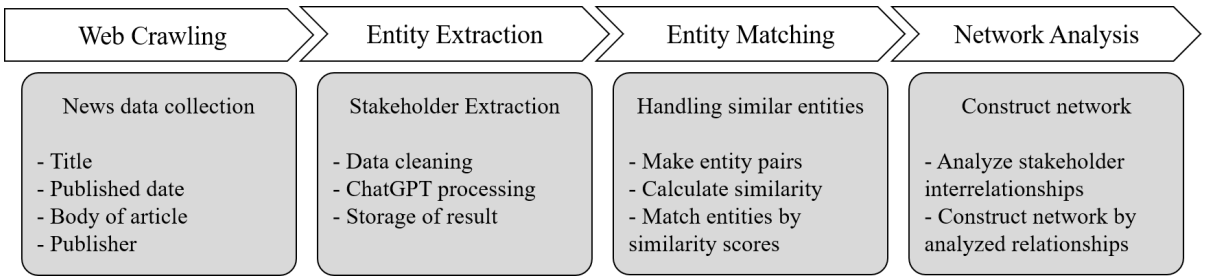

Figure 1. Stakeholder analysis process

3.1. Web Crawling

Web crawling is a technique to collect and store information automatically on web pages [13]. Web crawling has the characteristics of an exhaustive survey, and it enables one to obtain a large amount of data quickly and accurately with minimum human input [14]. News document published online is an important source of information that can be used for event detection and tracking [15]. Web crawling is an appropriate methodology to collect news data for a long period cost-effectively. Therefore, in this study, a web crawler was used to crawl a title, published date, publisher, and body of the news document published online.

3.2. Entity Extraction

Recent advancements in Large Language Models (LLMs) indicate their robust performance in diverse tasks, often without parameter tuning but merely with a few instructional examples [16]. Among these, ChatGPT [17], a derivative of the GPT pre-trained language model, stands out for its ability to comprehend user intent and produce human-like responses [18]. Remarkably, ChatGPT efficiently handles various NLP tasks with minimal input, avoiding the need for extensive fine-tuning [5], [6]. This study employs ChatGPT for Entity Extraction (EE), primarily due to its adeptness in understanding user intent, crucial for accurately identifying stakeholders in conflict-related projects. Additionally, ChatGPT's performance without extra fine-tuning, offers a cost-effective alternative to conventional language models that require extensive training datasets.

First, data cleaning process, which aims to remove unnecessary, irrelevant, or noisy information, is conducted to improve the quality of the text data and the performance of NLP models. Then, ChatGPT is utilized by system prompts and user prompts. A system prompt is a text to set the task framework for the model (specifying a role), and a user prompt is a text that a user inputs to guide the model's immediate response (giving a task). The system prompt, "You will be provided with an article, and your task is to extract a list of stakeholders involved in the conflict in relation to the infrastructure mentioned

in the article." was employed with two examples complementing as few-shot learning, and for the user prompt, a body of the article to analyze was entered. Results obtained from ChatGPT are finally stored in the form of lists of stakeholders.

3.3. Entity Matching

Entity Matching (EM) is a technique to identify representations that refer to the same real-world entity [19]. This study adopted a string similarity-based EM to handle various representations referring to the same entity. For example, expressions like "Governer Smith" and "Governer J.Smith" refer to the same real-world entity, but are recognized as different entities solely extracting entities by ChatGPT. The string similarity-based EM can effectively detect this kind of variation which is suitable for our analysis.

In the implementation of an EM process, similar entities within identified stakeholders were recognized and paired together. After removing duplicate entities, all possible entity pairs were generated and their similarity was calculated using Levenshtein ratio. Levenshtein ratio is a measure derived from Levenshtein distance, which quantifies the similarity between two strings by calculating the minimum number of single-character edits (insertions, deletions, or substitutions) required to change one sequence into the other [20]. Levenshtein ratio is defined as follows:

$$Levenshtein\ Ratio = \frac{(Length\ of\ the\ longer\ string - Levenshtein\ Distance)}{Length\ of\ the\ longer\ string} \quad (1)$$

A similarity threshold was set at 80%, above which entities were considered similar. Entity pairs that surpassed the similarity threshold were then grouped, and the results were systematically stored in the form of a dictionary.

3.4. Network Analysis

Network analysis is a tool to analyze the interdependencies among various elements [21]. Because of the advantage of being able to consider the interrelationship between elements well, it has begun to be widely used in the stakeholder study. Since the study of Mok et al. [11], it has begun to be used as the main direction of the stakeholder study [12]. Especially, in the context of social conflicts regarding the implementation of infrastructure projects, which are often complex, network analysis can offer valuable insights into the dynamics of these conflicts.

In this study, the stakeholder network was constructed by incorporating the extracted stakeholders on a year-by-year basis to understand the evolving dynamics of the conflict. Each node represents the stakeholders identified, and each link represents the interest in the same issue. The frequency of mention in the news article is based on the criticalness and impact of the stakeholders, and stakeholders mentioned in the same article are related stakeholders who are interested in the same issue. Based on this concept, the link was determined by the co-occurrence of stakeholders in the same article. For example, one link was connected when two stakeholders were mentioned in a single article. The weight of a link was assessed through the co-occurrence frequency of two stakeholders in one year.

For further analysis, node degree centrality was employed to identify critical stakeholders. Node degree centrality indicates how connected a stakeholder is to its immediate network neighbors, by calculating the total weight of the relationships that directly link to the node [22]. In this study, the calculation of nodal degree centrality in the stakeholder network is as follows:

$$C_{wd}(v) = \frac{\sum_{u \in N(v)} w(v,u)}{\sum_{x \in V} \sum_{y \in N(x)} w(x,y)} \quad (2)$$

Where $N(v)$ is a set of neighboring nodes directly connected to node v; $w(v,u)$ represents the weight of the edge connecting nodes v and u; V is the set of all nodes in the network; x and y are arbitrary nodes within the graph, with $N(x)$ being the set of neighboring nodes directly connected to node x.

4. Result

4.1. Case Selection

An experimental case study was employed to validate the proposed process. There were three major criteria for the case selection. First, the selected project should have been subject to varied social

discussions, indicative of social conflicts or debates. Second, the project needs to be relatively recent, ensuring that its development and progression can be comprehensively traced through collected online news documents. Last, the project should involve various stakeholders and encompass a variety of issues, reflecting its complexity and the dynamics of different interests and perspectives.

In accordance with these criteria, the Jeju 2nd Airport project was selected for the experiment. Jeju 2nd Airport project is a mega infrastructure project to construct a secondary international airport to alleviate the existing aviation demand pressures. Discussion on this project has been ongoing since the 1990s, and it has gradually materialized and is currently being promoted with the goal of opening the airport in 2026. The project is characterized by the involvement of a diverse range of stakeholders, with sharply polarized views on its pros and cons. Key issues such as local community awareness, environmental considerations, political dynamics, and demand concerns have emerged as focal points of discussion.

4.2. Data Collection

We utilized web crawling to collect data. Online news documents were crawled from the portal site Naver (www.naver.com, accessed on 30 January 2024). According to the digital news report by the Reuters Institute for the Study of Journalism, South Korea is one of the countries where online news consumption is very high, with a significant market share controlled by domestic platforms like Naver and Daum [23]. Consequently, the authors determined that incorporating news data from the portal site Naver is sufficient for the comprehensive analysis of stakeholder relationships in the context of Korean public infrastructure projects.

To see what discussions were made and what events occurred regarding Jeju 2nd Airport for 10 years, news documents from 2013 to 2022 were crawled with the input query of "Jeju 2nd Airport". As a result, a total of 11,651 news articles were collected. To facilitate the EE process using ChatGPT, data cleaning was conducted on each news article.

4.3. Experimental Result

There were 18,839 entities identified after eliminating exactly duplicate entities from 2013 to 2022. From the unduplicated list of entities, a list consisting of entity pairs was created and the similarity of each pair was calculated iteratively. Through this process, similarity value could be reviewed for all possible entity combinations, and a list of all possible similar entity pairs was derived. As a result, a total of 11,504 entities were identified to have the other representation of them, and these entities were categorized into 3,090 representative forms. Identified similar entities were then organized in the form of a dictionary.

Using the organized similar entities dictionary, for each news document's identified stakeholders, we iteratively reviewed if any of the identified stakeholders are in the similar entities dictionary. If the identified stakeholder is in the similar entities dictionary, it was replaced into its representative form regarding the dictionary. Through this process, a list of stakeholders for each news article that had been processed for similar entities was obtained. Then, under the regulations of nodes, links, and link weights mentioned previously, the stakeholder conflict networks were built annually from 2013 to 2022. The basic statistics of each network are presented in Table 1.

Table 1. Basic statistics of annual networks

Year	# of stakeholders	# of inter-stakeholder links
2013	14	6
2014	84	228
2015	1,427	8,189
2016	2,639	16,731
2017	1,805	12,215
2018	1,478	8,954
2019	2,326	15,476
2020	1,747	11,311
2021	2,158	18,067
2022	2,264	15,915

In analyzing a vast network with hundreds to thousands of nodes and edges, a subgraph comprising the top 100 nodes based on node degree was constructed to facilitate a more manageable and detailed analysis, recognizing the inherent significance of the larger network's structure. An example of a constructed network is presented in Figure 2. The presented network is the network of the year 2015.

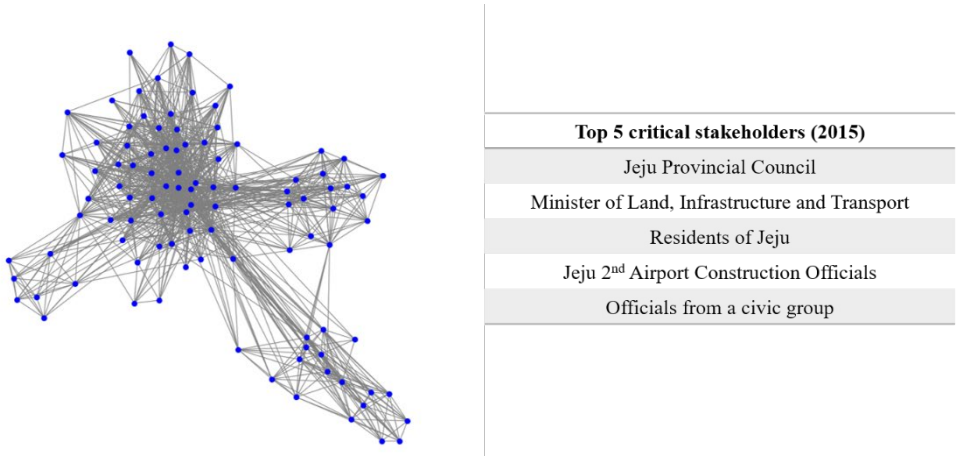

Figure 2. Stakeholder network with critical stakeholders on Jeju 2nd Airport in 2015

Based on this network, the top 5 critical stakeholders based on node degree centrality were identified. In 2015, the results of the feasibility study on Jeju Airport infrastructure expansion and development plan of Jeju 2nd Airport were announced. In the case of the feasibility study, it can be seen that government agencies such as the Ministry of Land, Infrastructure and Transport and the Jeju Provincial Council, which are major parties in this stage, have been identified as critical stakeholders. Also, civic groups, who are interested in the feasibility study results and development plan, have been also identified as critical stakeholders.

5. CONCLUSION

This study presents an innovative approach to identifying and analyzing stakeholders involved in social conflicts surrounding public infrastructure projects, utilizing ChatGPT in processing and analyzing textual data. By leveraging the news data and employing entity extraction, entity matching, and network analysis, this research systematically identifies stakeholders and elucidates the complex interrelations and dynamics among them. The case study on the Jeju 2nd Airport project explores the effectiveness of our proposed process, demonstrating its potential to enhance the understanding and management of stakeholder dynamics in public infrastructure projects. This study advances the field by providing an efficient methodology for stakeholders in the context of social conflicts over public infrastructure projects. Leveraging NLP technology enhances our understanding of complex stakeholder landscapes in a timely manner, laying crucial groundwork for advancing conflict analysis and stakeholder management research.

Despite its contributions, this study has limitations that warrant future research. First, the analysis was confined to node degree centrality within the network, indicating a narrower focus than possible with more comprehensive and nuanced network analyses. The authors are planning to refine and deepen this analytical approach. Second, the construction of the stakeholder network relied solely on stakeholder data, omitting conflict drivers and stakeholder interests. Incorporating these elements could significantly enhance the network's complexity and the depth of research achievable, suggesting a more room for elaborate future research.

ACKNOWLEGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2022R1A2C1012018).

REFERENCES

[1] J. Eweje, R. Turner, and R. Müller, “Maximizing strategic value from megaprojects: The influence of information-feed on decision-making by the project manager”, *Int. J. Proj. Manag.*, vol. 30, no. 6, pp. 639–651, 2012, doi: 10.1016/j.ijproman.2012.01.004.

[2] J. Ninan, A. Mahalingam, and S. Clegg, “Power in news media: Framing strategies and effects in infrastructure projects”, *Int. J. Proj. Manag.*, vol. 40, no. 1, pp. 28–39, 2022, doi: 10.1016/j.ijproman.2021.09.003.

[3] D. Namgoong, J. Lee, S. Baek, W. J. Woo, C. Lee, and S. H. Han, “Sentiment Analysis Model For Public Construction Projects Using KoBERT”, *Proc. Creative Constr. Conf. 2023*, pp. 359–366, 2023, doi: 10.3311/ccc2023-048.

[4] C. Lee, J. W. Won, W. Jang, W. Jung, S. H. Han, and Y. H. Kwak, “Social conflict management framework for project viability: Case studies from Korean megaprojects”, *Int. J. Proj. Manag.*, vol. 35, no. 8, pp. 1683–1696, 2017, doi: 10.1016/j.ijproman.2017.07.011.

[5] M. T. R. Laskar, M. S. Bari, M. Rahman, M. A. H. Bhuiyan, S. Joty, and J. X. Huang, “A Systematic Study and Comprehensive Evaluation of ChatGPT on Benchmark Datasets”, *arXiv*, 2023, doi: 10.48550/arxiv.2305.18486.

[6] B. Lamichhane, “Evaluation of ChatGPT for NLP-based Mental Health Applications”, *arXiv*, 2023, doi: 10.48550/arxiv.2303.15727.

[7] S. Olander, “Stakeholder impact analysis in construction project management”, *Constr. Manag. Econ.*, vol. 25, no. 3, pp. 277–287, 2007, doi: 10.1080/01446190600879125.

[8] A. A. Elias, “A system dynamics model for stakeholder analysis in environmental conflicts”, *J. Environ. Plan. Manag.*, vol. 55, no. 3, pp. 387–406, 2012, doi: 10.1080/09640568.2011.604191.

[9] T. H. Y. Li, S. T. Ng, and M. Skitmore, “Evaluating stakeholder satisfaction during public participation in major infrastructure and construction projects: A fuzzy approach”, *Autom. Constr.*, vol. 29, pp. 123–135, 2013, doi: 10.1016/j.autcon.2012.09.007.

[10] J. Lienert, F. Schnetzer, and K. Ingold, “Stakeholder analysis combined with social network analysis provides fine-grained insights into water infrastructure planning processes”, *J. Environ. Manag.*, vol. 125, pp. 134–148, 2013, doi: 10.1016/j.jenvman.2013.03.052.

[11] K. Y. Mok, G. Q. Shen, and R. J. Yang, “A network theory-based analysis of stakeholder issues and their interrelationships in large construction projects: a case study”, *Int. J. Constr. Manag.*, vol. 17, no. 3, pp. 210–227, 2017, doi: 10.1080/15623599.2016.1187246.

[12] J. Xue, G. Q. Shen, R. J. Yang, I. Zafar, and E. M. A. C. Ekanayake, “Dynamic Network Analysis of Stakeholder Conflicts in Megaprojects: Sixteen-Year Case of Hong Kong-Zhuhai-Macao Bridge”, *J. Constr. Eng. Manag.*, vol. 146, no. 9, 2020, doi: 10.1061/(asce)co.1943-7862.0001895.

[13] M. Kovacevic, J.-Y. Nie, and C. Davidson, “Providing Answers to Questions from Automatically Collected Web Pages for Intelligent Decision Making in the Construction Sector”, *J. Comput. Civ. Eng.*, vol. 22, no. 1, pp. 3–13, 2008, doi: 10.1061/(asce)0887-3801(2008)22:1(3).

[14] E. Ferrara, P. D. Meo, G. Fiumara, and R. Baumgartner, “Web data extraction, applications and techniques: A survey”, *Knowl.-Based Syst.*, vol. 70, pp. 301–323, 2014, doi: 10.1016/j.knosys.2014.07.007.

[15] I. Mele, S. A. Bahrainian, and F. Crestani, “Event mining and timeliness analysis from heterogeneous news streams”, *Inf. Process. Manag.*, vol. 56, no. 3, pp. 969–993, 2019, doi: 10.1016/j.ipm.2019.02.003.

[16] X. Wei, X. Cui, N. Cheng, X. Wang, X. Zhang, S, Huang, P. Xie, J. Xu, Y. Chen, M. Zhang, Y. Jiang, W. Han, “Zero-Shot Information Extraction via Chatting with ChatGPT”, arXiv, 2023, doi: 10.48550/arxiv.2302.10205.

[17] L. Ouyang, J. Wu, X, Jiang, D. Almeida, C.L. Wainwright, P. Mishkin, C.Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F.Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, R. Lowe, “Training language models to follow instructions with human feedback”, arXiv, 2022, doi: 10.48550/arxiv.2203.02155.

[18] B. Li, G. Fang, Y. Yang, Q. Wang, W. Ye, W. Zhao, S. Zhang, “Evaluating ChatGPT's Information Extraction Capabilities: An Assessment of Performance, Explainability, Calibration, and Faithfulness”, arXiv preprint arXiv:2304.11633.

[19] N. Barlaug and J. A. Gulla, “Neural Networks for Entity Matching: A Survey”, *ACM Trans. Knowl. Discov. Data (TKDD)*, vol. 15, no. 3, pp. 1–37, 2021, doi: 10.1145/3442200.

[20] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, “Duplicate Record Detection: A Survey”, *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 1, pp. 1–16, 2007, doi: 10.1109/tkde.2007.250581.

[21] S. Wasserman and K. Faust, “*Social Network Analysis: Methods and Applications”,* Cambridge: Cambridge University Press, 1994.

[22] M. Loosemore, “Social network analysis: using a quantitative tool within an interpretative context to explore the management of construction crises”, *Eng. Constr. Arch. Manag.*, vol. 5, no. 4, pp. 315–326, 1998, doi: 10.1046/j.1365-232x.1998.54039.x.

[23] E. Newman, R. Fletcher, D. A. L. Levy, and R. K. Nielsen, “Reuters Institute Digital News Report 2018”, Reuters Institute for the Study of Journalism, University of Oxford, Oxford, U.K., Tech. Rep., 2018. [Online]. Available: http://www.digitalnewsreport.org/