Automatic Inference of Construction Delays through Analysis of Weekly Progress Reports Using LLMs

Aritra Pal (corresponding author), (ap2425@cam.ac.uk)
Department of Engineering, University of Cambridge, United Kingdom

Danny Murguia, (dem52@cam.ac.uk)
Department of Engineering, University of Cambridge, United Kingdom

Campbell Middleton, (crm11@cam.ac.uk)
Department of Engineering, University of Cambridge, United Kingdom

Keywords: Delays, construction progress, weekly reports, natural language processing (NLP), large language models (LLM),

Abstract

Construction projects frequently face delays, presenting significant challenges to effective project management. Identifying the causes of these delays and determining which stakeholders are responsible for addressing them is crucial for effective mitigation. While progress reports from construction sites offer valuable insights into project status, they are primarily used for contract management. Consequently, despite containing crucial data on project delays, they are often underutilised for delay management. Moreover, manual comprehension and analysis of these reports further complicate the identification of delay causes, hindering thorough analysis. To address this issue, this research proposes an automated delay inference method using natural language processing (NLP) techniques. By leveraging large language models (LLMs), the aim is to extract critical project information, including delay causes, from 44 weekly progress reports. These causes are then categorised based on their location and nature. Finally, a flowline diagram is generated to visualise the planned and actual construction programmes. The gaps between them are explained by the identified delay causes and their impact on project completion, providing instant insights for effective decision-making. This approach aims to enhance the accuracy and efficiency of delay analysis in construction projects, ultimately improving project management.

1. Introduction

Delays are pervasive in the construction industry, leading to cost overruns, contractual disputes, and compromised project quality (Park, 2021). Accurate and timely identification of the cause of delay is essential for effective project management and mitigation strategies. Weekly progress reports (WPRs) from construction sites provide valuable insights into a project's performance, documenting progress and issues that disrupt construction progress. However, since these reports are written by site personnel in natural language, they often lack structure. Traditionally, extracting delay-related information from WPRs requires manual reading and comprehension, which is labour-intensive and prone to errors (Fertitta et al., 2016). Additionally, identifying the causes of delays is not always straightforward, particularly in cases of contractor inefficiency (Kao & Yang, 2009). While delays caused by total work stoppages, such as labour strikes, are clear-cut, attributing delays to contractor inefficiencies, complexities are further compounded where liability attribution varies based on the complexities of each situation (Keane & Caletka, 2015). So, classification of the issues causing delays early on in the project and keeping track of those delays is necessary for mitigating them in a timely manner. Recent advancements in natural language processing (NLP) and machine learning offer promising avenues for automating this process (Ye et al., 2023), thereby enhancing the efficiency and accuracy of delay identification and classification from WPRs.

This paper proposes a novel methodology for automatic inference of construction delays using advanced NLP techniques. By leveraging OpenAI's large language model (LLM), this study aims to extract critical information from WPRs, classify delay-causing issues, and visualise their impacts on construction schedules. The methodology encompasses several steps: extracting and structuring data from WPRs, classification of delay causes, flowline representation of construction programmes, and visualisation of delay impacts. To validate the effectiveness of this approach, it was applied to a high-rise building construction project in London, analysing 44 WPRs over 44 weeks. The results demonstrate significant improvements in the accuracy of delay identification and classification and enhanced clarity in visualising the impact of delays on the project timeline. This study underscores the potential of integrating NLP and machine learning techniques into construction management practices, paving the way for more data-driven and responsive project management solutions. The following sections provide a detailed account of the related studies, the methodology, the case study, and the results obtained, along with a discussion of the implications for construction project management.

2. Related studies

This section reviews related studies on delay analysis in construction and the application of artificial intelligence in delay analysis.

2.1. Delay analysis in construction projects

Delay analysis in construction is a crucial and complex process involving identifying, assessing, and managing delays that can significantly impact project completion and costs. The delay analysis process, typically conducted by expert consultants, involves three stages: information retrieval, delay analysis, and communication of findings (Boyacioglu et al., 2022). However, each stage presents challenges, with information retrieval being particularly time-consuming and costly due to the construction industry's poor documentation practices. Notably, it consumes about 70% of the delay analysis effort (Ali et al., 2020). Effective delay analysis requires a robust method to model the causal relationships between delay events and project completion time (Axelson, 2021). Despite efforts by organisations like the Society of Construction Law and the American Association of Cost Engineering to standardise these methods, their efficacy is still heavily reliant on the quality and availability of supporting information. Regular and timely analysis of WPRs can be one of the solutions. However, the traditional method of manual analysis shall be replaced with emerging technologies (Boyacioglu et al., 2022).

2.2. Artificial intelligence (AI) in delay analysis

AI is revolutionising data analysis in construction by significantly enhancing its efficiency and precision (Pal & Hsieh, 2021). AI and ML facilitate the processing and interpretation of vast amounts of data, addressing major challenges such as information retrieval and the clear representation of analysis. AI systems can automate the identification and sourcing of relevant information. Furthermore, ML algorithms can model the causal relationships between delay events and project timelines, providing a more accurate and reliable analysis compared to traditional methods (Soibelman & Kim, 2002). NLP, a subfield of AI, helps extract meaningful insights from unstructured textual data such as project reports, schedules, and contracts, which are often rich sources of information in construction projects (Pal et al., 2023). NLP can automatically classify and summarise large volumes of text, highlight potential issues, and track the evolution of a project over time (ul Hassan et al., 2021). AI's ability to learn from data and improve over time makes it an invaluable tool for delay analysts, enabling more effective identification, analysis, and communication of delay causes and their impacts (Egwim et al., 2021). Although NLP has proved to be an effective method in construction data analysis, its application in delay inference is still very limited (Gondia et al., 2020)

3. Methodology

Weekly progress reports originating from construction sites are a good source of valuable information on project performance. Along with the progress status, it also documents issues that hinder the smooth occurrence of construction tasks at the site. As site managers write these descriptive reports in natural language, they are often unstructured. Also, the only means of extracting information related to project delays from them is manual reading and comprehension. Considering the scale of a project, it is often difficult to interpret and keep track of all the issues reported through WPRs. With the advancement of natural language processing capabilities in recent times, computer-aided analysis of WPRs can be a way forward for automatic inference of schedule slippage during construction.

The following method is proposed to address this challenge. A graphical representation of the proposed delay inference approach is shown in Figure 1. It is divided into four steps: extracting data from weekly progress reports and representing them in a structured format, identifying issues causing delays during construction and classifying them as per pre-defined delay cause classes, representing baseline and updated programmes through flowlines, and visualisation of the impact of delay causes on the project till date.

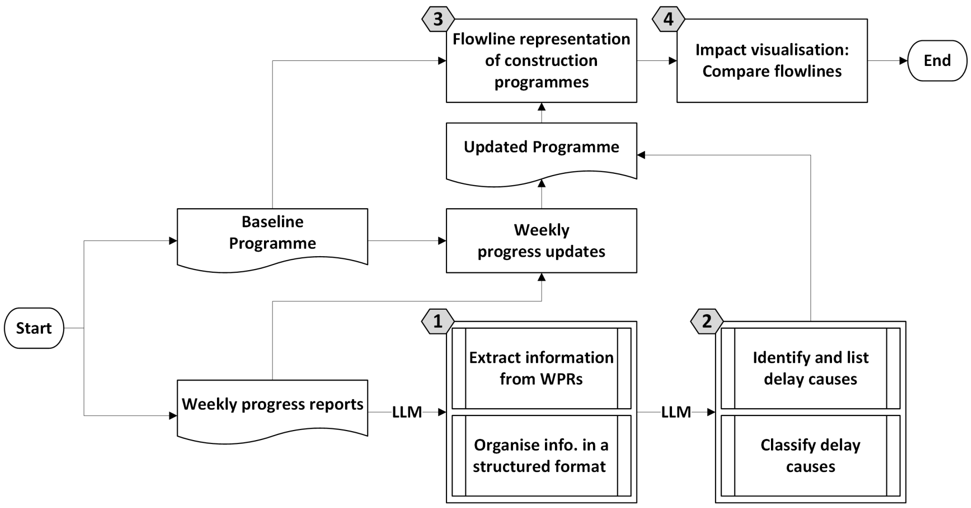

Figure 1: Methodology diagram

3.1. Information extraction from weekly progress reports

This step utilises OpenAI's 'gpt-3.5-turbo-0125' large language model to extract critical information from Weekly Progress Reports (WPRs) in Word format. The model is employed to identify and retrieve essential details such as the report's author, reporting date, the specific week covered by the report, any delays occurring within that week, cumulative delays up to that point, the baseline programme used for delay estimation, and issues related to various project structures. This extraction is achieved through prompt engineering, where the model is guided by meticulously crafted prompts detailing each category of information. The extracted data is then systematically structured and stored in a JSON file for subsequent analysis and reference.

3.2. Identification and classification of delay-causing issues

Identifying and classifying delay-causing issues from WPRs are also conducted using OpenAI's 'gpt-3.5-turbo-0125' LLM. This process entails extracting critical information from the previously stored JSON file, specifically focusing on issues related to various project work packages such as foundation works, structural frames, and mechanical installations. Subsequently, issues related to each component undergo further analysis, and each issue is classified according to predefined delay cause classes. Prompt engineering guides the model with detailed descriptions of each delay cause category, enabling accurate identification and retrieval of data on factors contributing to delays. These factors encompass material shortages, labour constraints, design modifications, planning and scheduling issues, etc. Table 1 presents a comprehensive list of delay causes and their details. These details are supplied to the LLM as a system prompt and delay-causing issues as user queries. The LLM is instructed to output results in a structured JSON format.

3.3. Flowline representation of construction programmes

In this step, flowlines are selected to visualise the construction programmes with a location breakdown structure on a level-by-level basis. Flowlines provide construction stakeholders with a better tool for comprehending the construction programme. It helps illustrate the relationships and interdependencies of timeframes between consecutive levels more effectively (Murguia et al., 2023). A Python wrapper of the Java library 'MPXJ' is used to extract level-wise start dates and finish dates and populate them to flowlines automatically. MPXJ enables the reading of construction project programmes from various file formats and databases, such as Asta Power Projects, Primavera P6, Microsoft Project, and so on. Level-wise baseline start and finish dates are used to construct planned flowlines where actual dates create flowlines that represent actual progress.

3.4. Delay impact visualisation on flowlines

Visualising delay impacts on flowlines provides a comprehensive and intuitive representation of how delays affect the construction programme. By overlaying delay causes onto the flowlines, stakeholders can quickly grasp the extent and implications of schedule disruptions at different project stages. Each delay cause is visually depicted along the floor-wise flowline, offering insights into its magnitude, duration, and specific location within the project timeline. This visualisation allows project managers to identify bottlenecks and potential ripple effects of delay causes across various project components. Additionally, by comparing the planned flowlines with the actual progress incorporating delays, stakeholders can better understand schedule deviations and make informed decisions to mitigate future delays.

4. Case study

The proposed methodology was tested on a commercial building construction project in London. This project comprises an office space with two towers, one with nine floors (Building A) and the other with eight (Building B), covering a total area of 139,000 square

Table 1: Details of delay causes

Delay causes	Details
Labour	Labour shortage; Lack of labour competence; Lack of labour motivation
Materials	Material shortage; Material delivery issues; Material not available in the work area; Material quality issues
Equipment	Equipment in use by other trades; Equipment not available in the work area; Equipment breakdown; Equipment in maintenance
Planning and Scheduling	Poor sequencing of activities; Late handover from previous trade; Other trades in the work area; Work area not available
Site logistics	Loading-bay areas unavailable; Temporary storage far from work area; Site facilities far from work area
Health and safety	Labour being trained/inducted; Safety non-compliance; Accident; Improper health and safety plan; RAMS not in place; Permit not in place; Drug testing
Design and constructability	Design information not available; Design issues; Buildability/ Constructability
Quality	Re-work; Quality control issues
Change	Change orders - scope change
Weather	Strong winds; Heavy rain; Temperature

Adapted from (Rathnayake & Middleton, 2023).

feet. The estimated duration of the project was 88 weeks. The actual construction of the structural frame for the two towers occurred between weeks 38 and 84. This study analysed 44 WPRs submitted during this period. Each report was written in a descriptive Word format and was 10-12 pages long. The project followed two baseline programmes, with weekly delays or accelerations recorded against these baselines. A revised baseline programme was introduced in week 60. This study employed an automatic delay inference approach to identify delays, specifically in the structural frame construction of both towers.

4.1. Results and Discussions

Following the first step of the methodology, cumulative delays and weekly delays for the entire project at the end of each week were extracted from the WPRs. The cumulative delay results are shown in Figure 2. It is observed that by week 59, the project delay had accumulated to 87 days, highlighting the need for a revised baseline program.
Following the extraction of delay durations from the WPRs, delay-causing issues for buildings A and B were identified and listed according to step 2 of the methodology. A total of 562 issues were identified during the construction of buildings A and B. Then, each issue was systematically classified into one of ten predefined categories from Table 1. Accurate categorisation was ensured by meticulously analysing the context and details of each delay, providing a clear and structured overview of the root causes impacting project completion. This classification helped identify patterns and areas needing improvement, facilitating more effective project management and mitigation strategies.

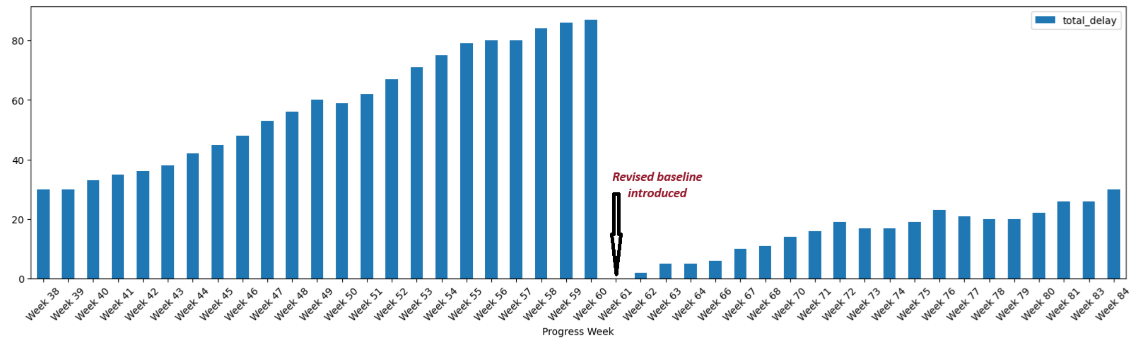

Figure 2: Cumulative number of days delayed against baseline programme

A validation set of 157 issues was created to evaluate the performance of the delay-cause classification model. The validation set issues were manually classified to compare with the predicted classification results. The classification results were evaluated using three standard metrics: precision (P), recall (R), and F1 score. Precision measures the accuracy of the positive predictions made by the model. It is the ratio of true positive (TP) predictions to the total number of positive predictions (true positives + false positives). Recall, also known as sensitivity or true positive rate, measures the model's ability to identify all relevant instances. It is the ratio of true positive predictions to the total number of actual positive instances (true positives + false negatives). The F1 score is the harmonic mean of precision and recall, providing a single metric that balances both concerns. It is particularly useful when the class distribution is imbalanced. The F1 score ranges from 0 to 1, with 1 being the best possible score, indicating perfect precision and recall. P, R, and F1 values were estimated for each delay class, and the average value was estimated by taking a mean of all classes. Table 2 shows the average P, R, and F1 values for the delay cause classification model. High values indicate considerable good performance of the LLM in classifying delay causes. A confusion matrix was created to understand the performance of the multiclass delay cause classification. In multi-class classification, the confusion matrix is an extension of the binary confusion matrix, where each row represents the instances in an actual class, and each column represents the instances in a predicted class. This matrix helps evaluate a classification model's performance across multiple classes. The confusion matrix for the delay cause classification is shown in Figure 3. Row normalisation was applied to this matrix. It normalises the counts in each row by the total number of actual instances in the corresponding class. This gives the proportion of predictions for each class relative to the true instances of that class. Normalisation makes it easier to interpret the confusion matrix, especially when dealing with large or imbalanced datasets. Diagonal Elements refer to the correct predictions, whereas off-diagonal elements refer to misclassifications. Value 1 indicates a 100% accurate prediction of all class instances. In this study, a few instances of the 'labour' class were misclassified as 'Change', and a few instances of the quality class were misclassified as labour and materials. This indicates that the issue description written in the WPR must be clearly stated. Also, it highlights the requirements for clearer instructions to be promoted to the LLM to help the model classify complex issues.

Following the successful evaluation of the LLM, it was utilised to classify all issues related to buildings A and B, derived from WPRs. A few examples are shown in Figure 4. The duration of delays extracted from each WPR was then mapped to the delay causes during that week. This mapping helped establish clear cause-and-effect relationships between delay causes and the number of days the project was impacted. The impact

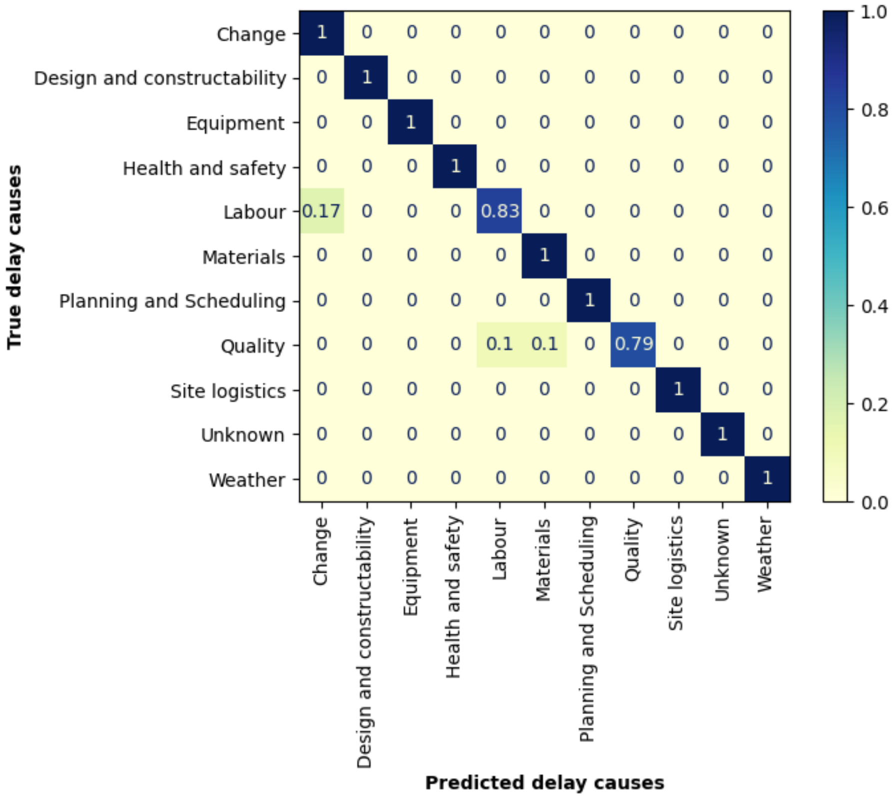

Figure 3: Confusion matrix

```
"block_A_issues": [
    {
        "description": "No laydown areas on site for structural steels",
        "delay_cause_class": "Site logistics"
    },
    {
        "description": "Out of sequence steel deliveries",
        "delay_cause_class": "Planning and Scheduling"
    },
```

Figure 4: Examples of delay cause classification

of each delay category was estimated proportionately based on the occurrence of each delay cause. For example, if the total delay in a week was four days and three out of five identified issues were related to 'Planning and scheduling', the delay attributed to 'Planning and scheduling' would be calculated as (3/5)×4 days, equating to 2.4 days. The weekly delays over the course of 44 weeks were plotted and categorised by delay causes. Figure 5 displays the same. It is observed that the delay in a single week could reach up to 5 days, such as weeks 47 and 52.

Further analysis was conducted to determine the proportion of each delay category in the overall project delay. Figure 6 shows a pie chart representing this distribution. It is observed that 'Planning and Scheduling' accounted for the maximum delays during the construction of building A, with a 25% proportion. 'Materials', 'quality', and 'labour' were identified as the next major causes of delay. Understanding the distribution of delay causes helps project managers prioritise areas for improvement, allocate resources more effectively, and implement targeted strategies to mitigate future delays. By addressing the primary sources of delays, the project team can enhance efficiency and reduce the impact on the overall project timeline.

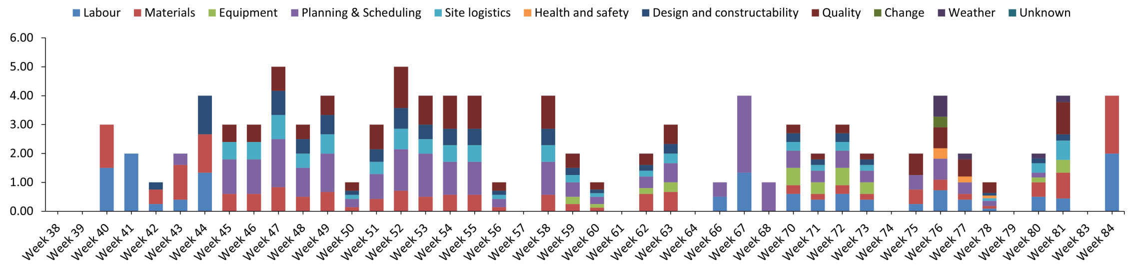

Figure 5: Number of days delayed per week with causes

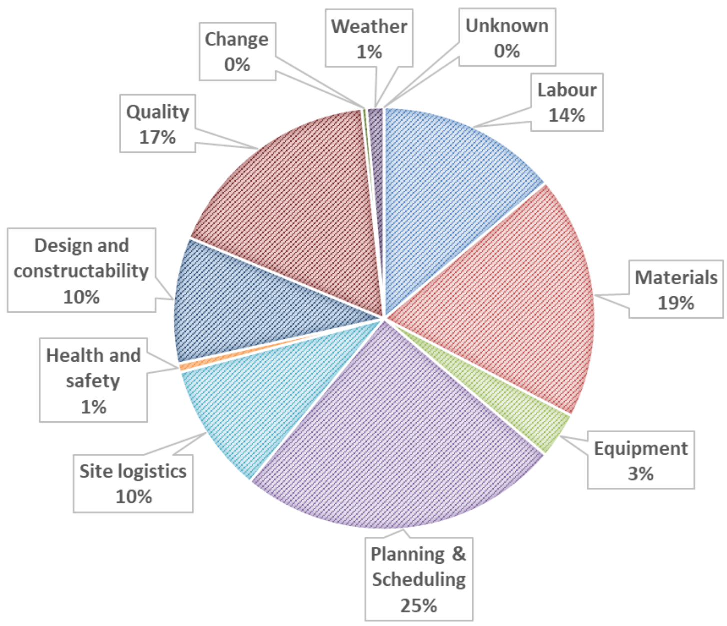

Figure 6: Proportion of different delay causes

Planned and actual start and finish dates for buildings A and B were extracted from the baseline and updated programs as described in step three of the methodology. Flowlines were constructed for each level of the buildings. By comparing the actual completion date of each level with the week in which delays occurred, delay causes were associated with specific building levels. The cumulative total delay at the end of each level of construction, along with the related delay causes, was then displayed on the flowline. This visualisation could give stakeholders an immediate understanding of the project status, deviations from the planned programme, and major causes of delays. By offering a clear and comprehensive view of the project's progress and pinpointing critical issues, this method enhances project management efficiency by enabling more informed decision-making. As an example, Figure 7 shows the flowline representation of building A along

Table 2: Performance of delay cause classification model

Metrics	Average value
Precision	0.91
Recall	0.97
F1 Score	0.93

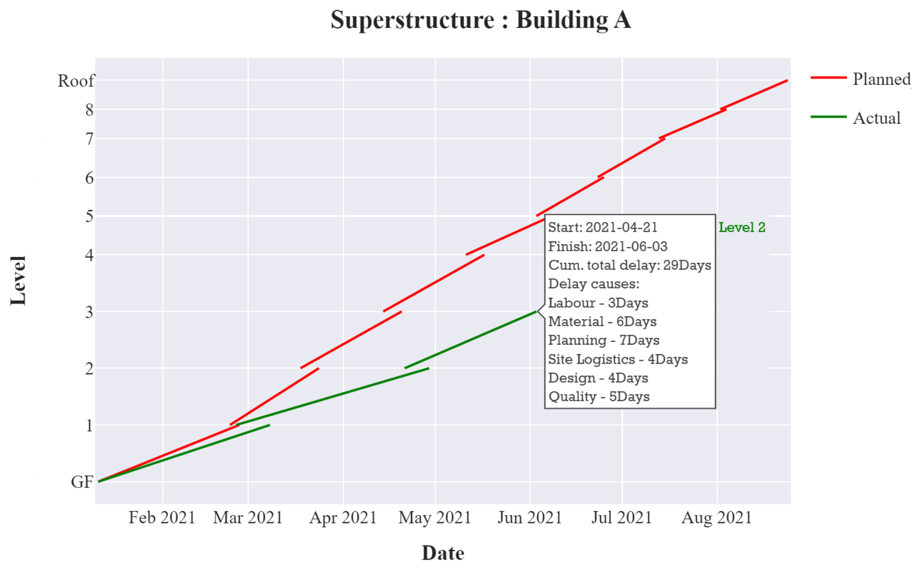

Figure 7: Visualisation of delay causes overlaid on flowlines

with the identified delay causes up to level 3. This figure clearly indicates that the cumulative total delay until the completion of level 3 was 29 days. The delay was attributed to various factors: labour accounting for three days, materials for six days, planning and scheduling for seven days, site logistics for four days, design and constructability for four days, and quality issues for five days.

5. Conclusion

This study introduces an automatic approach for inferring construction delays from weekly progress reports (WPRs). It demonstrates the effectiveness of using natural language processing techniques to enhance the accuracy and efficiency of delay analysis in construction projects. By leveraging large language models (LLMs), critical project information, including delay causes, is successfully extracted from WPRs. The categorisation of these causes based on their location and nature, followed by the generation of flowline diagrams, provided clear visual insights into the impact of delays on project completion. This automated approach addresses the limitations of manual analysis of WPRs, reducing the risk of oversight and inaccuracies in delay-cause identification and offering instant insights for effective decision-making. By improving the identification and analysis of delay causes, this method supports better resource allocation, targeted mitigation strategies, and overall project management, leading to more timely and cost-effective project completions. Future research in this direction shall explore the usage of open-source LLMs and fine-tuning them with construction delay-specific datasets.

Acknowledgements

We would like to thank our industry partners for providing us with access to actual construction data, their engagement with this research, and the visionary approach to data-driven decision-making.

References

Ali, B., Zahoor, H., Nasir, A. R., Maqsoom, A., Khan, R. W. A., & Mazher, K. M. (2020). Bim-based claims management system: A centralized information repository for extension of time claims. *Automation in Construction*, *110*, 102937. https://doi.org/https://doi.org/10.1016/j.autcon.2019.102937

Axelson, A. S. (2021). *Delay analysis in construction contracts*. The society of construction law (UK).

Boyacioglu, S. E., Greenwood, D., & Rogage, K. (2022). Incorporating emerging technologies in the forensic analysis of construction project delays. *IOP Conference Series: Earth and Environmental Science*, *1101*(5), 052029. https://doi.org/10.1088/1755-1315/1101/5/052029

Egwim, C. N., Alaka, H., Toriola-Coker, L. O., Balogun, H., & Sunmola, F. (2021). Applied artificial intelligence for predicting construction projects delay. *Machine Learning with Applications*, *6*, 100166. https://doi.org/https://doi.org/10.1016/j.mlwa.2021.100166

Fertitta, T. D., Nedinsky, A. L., & Gilmore, J. G. (2016). Construction project delays and time extensions. In *Construction contract claims, changes, and dispute resolution* (pp. 117–137). https://doi.org/10.1061/9780784414293.ch05

Gondia, A., Siam, A., El-Dakhakhni, W., & Nassar, A. H. (2020). Machine learning algorithms for construction projects delay risk prediction. *Journal of Construction Engineering and Management*, *146*(1), 04019085. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001736

Kao, C.-K., & Yang, J.-B. (2009). Comparison of windows-based delay analysis methods. *International Journal of Project Management*, *27*(4), 408–418. https://doi.org/https://doi.org/10.1016/j.ijproman.2008.05.016

Keane, P. J., & Caletka, A. F. (2015). *Delay analysis in construction contracts*. John Wiley; Sons.

Murguia, D., Rathnayake, A., & Middleton, C. Master schedule optimisation with the use of flowlines and performance data. English. In: In *Proceedings of the 31st annual conference of the international group for lean construction (iglc31)*. Lille, France, 2023, 1463–1474. https://doi.org/10.24928/2023/0121

Pal, A., & Hsieh, S.-H. (2021). Deep-learning-based visual data analytics for smart construction management. *Automation in Construction*, *131*, 103892. https://doi.org/https://doi.org/10.1016/j.autcon.2021.103892

Pal, A., Lin, J. J., & Hsieh, S.-H. (2023). Schedule-driven analytics of 3d point clouds for automated construction progress monitoring. In *Computing in civil engineering 2023* (pp. 412–420). https://doi.org/10.1061/9780784485224.050

Park, J. E. (2021). Schedule delays of major projects: What should we do about it? *Transport Reviews*, *41*(6), 814–832. https://doi.org/https://doi.org/10.1080/01441647.2021.1915897

Rathnayake, A., & Middleton, C. (2023). Systematic review of the literature on construction productivity. *Journal of Construction Engineering and Management*, *149*(6), 03123005. https://doi.org/10.1061/JCEMD4.COENG-13045

Soibelman, L., & Kim, H. (2002). Data preparation process for construction knowledge generation through knowledge discovery in databases. *Journal of Computing in Civil Engineering*, *16*(1), 39–48. https://doi.org/10.1061/(ASCE)0887-3801(2002)16:1(39)

ul Hassan, F., Le, T., & Lv, X. (2021). Addressing legal and contractual matters in construction using natural language processing: A critical review. *Journal of Construction Engineering and Management*, *147*(9), 03121004. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002122

Ye, Y.-X., Shan, M., Gao, X., Li, Q., & Zhang, H. (2023). Examining causes of disputes in subcontracting litigation cases using text mining and natural language processing techniques. *International Journal of Construction Management*, *0*(0), 1–13. https://doi.org/10.1080/15623599.2023.2286046