EVALUATING THE COMPREHENSION OF CONSTRUCTION SCHEDULES OF AN ARTIFICIAL INTELLIGENCE

Tulio Sulbaran, Ph.D.
The University of Texas at San Antonio, Texas, United States.

ABSTRACT: *Construction schedules are an important tool to communicate with the project stakeholders and are critical for the project management team to plan, coordinate, and manage construction projects. Each construction project has a unique schedule that is created based on the construction drawings, specifications, contracting requirements, construction methods, and the judgment of the project management team. Therefore, each construction schedule is unique in many aspects such as the number of activities, the names of the activities, the duration of those activities, and the relationship between the activities. The names of the activities are of particular interest as they are the critical core unit to creating the schedule. Furthermore, the activities are the ones that bring together all other aspects of the schedule. Unfortunately, there is no standard naming conversion for those activities and they vary from project to project as well as from project management team to project management team. This inconsistency of the activity name makes it extremely challenging for both humans and machines to understand the meaning and scope of the activities. Thus, the problem that this paper addresses is the challenge faced by machines to comprehend the activities of a construction schedule. Therefore, the objective of this paper is to evaluate the ability of an Artificial Intelligence (AI) implementation to comprehend activities in a construction schedule. This research was conducted following a mixed research method. The AI implementation training was done by providing the Construction Specifications Institute (CSI) Master Format activity list to a Sentence Transformer. Then the AI was given the task of interpreting the activities of a construction schedule according to the 50 Divisions of the CSI Master Format. A group of senior construction students was also given the same interpretation task. The evaluation was done by comparing the results of the AI vs the humans for each of the activities in the construction schedule. The result was that the AI has 0.56 accuracy, 0.50 precision, 0.85 recall and, 0.64 F1 Score. This result is very promising and it supports further research to refine the AI to increase its ability to comprehend construction schedule activities. Upon achieving a higher level of comprehension an AI could be used to assist humans in the preparation of construction schedules or perhaps prepare drafts of the construction schedules for the human to review.*

KEYWORDS: *Construction Scheduling, Decision Support, Artificial Intelligence, Comprehension*

1. BACKGROUND

Construction scheduling is a complex process with a lot of considerations for successful project delivery with different and specific approaches for each scheduling constraint (Okonkwo et al., 2022). Preparing good schedules is a time-consuming process that requires a deep understanding of the construction process (Sulbaran, & Ahmed, 2017). Construction schedules serve many purposes ranging from informing owners on state of progress, establishing long-term coordination among crews and trade contractors, to specifying terms of payment (Halpin & Senior, Bolivar, 2017). The construction schedule is one of the most important planning and control tools for the construction process (Rosłon et al., 2020), frequently includes a very large number of activities (Essam et al., 2023) and it is the core of the project plan. It is used by the project management team to commit resources to the project and show the organization how the work will be performed (Magalhães-Mendes, 2011). The main goal of a construction schedule is to identify the activities needed to complete a project and sequence them in the most efficient way possible within the timeframe and resources available (Essam et al., 2023). Construction scheduling is a complex process due to the interdependence and contradiction of project activities (Essam et al., 2023). Construction schedule practices rely heavily on manually elaborated descriptions of construction means and methods (Amer & Golparvar-Fard, 2019). The preparation of a construction schedule including the number of activities, the names of the activities, the duration of those activities, and the relationship between the activities which heavily relies on the judgment and expertise of the project management team.

The names of the construction activities are the only unstructured data attribute in the construction schedules (Hong et al., 2021). Construction activities are described using Natural Language expressions with little or no standardization, grammatical errors, abbreviations, project and construction-specific terms (Heigermoser et al., 2019). Construction activities have been widely discussed in the construction literature (Amer & Golparvar-Fard, 2019) as they are critical in construction schedules. The activity names are devised to communicate between stakeholders, however, they are often written using inconsistent terminologies with omitted contextual information

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)
Tulio Sulbaran, *Evaluating the Comprehension of Construction Schedules of an Artificial Intelligence*, pp. 545-552, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.53

(Hong et al., 2021). The inconsistency and omissions are due in part because construction schedules are prepared by project management teams' using their tacit knowledge. The tacit knowledge is the common knowledge on the process of conformance checking that is applied by domain experts (Yurchyshyna & Zarli, 2009). This inconsistency in the activity names is further aggravated by the variety of construction means and methods to perform construction activities and the differences in practice between different construction companies (Amer & Golparvar-Fard, 2019). It is also the case that in most instances, historic information including scheduling decision reasoning is not documented and disseminated for use in other future projects (Hong et al., 2022). Although construction companies might establish procedures to propagate their construction scheduling knowledge between different projects and teams (Amer & Golparvar-Fard, 2019), it is ultimately the project management team that prepares the construction schedule. This current scheduling practice leads to activities written in an inconsistent format with inconsistent terminologies (Hong et al., 2021) which makes it extremely challenging for both humans and machines to understand the meaning and scope of the activities.

The problem addressed by this paper is the challenge faced by machines to comprehend the activities of a construction schedule. Thus, the objective of this paper is to evaluate the ability of an Artificial Intelligence (AI) implementation to comprehend activities of a construction schedule. The AI's ability to comprehend construction activities is critical to further advance the AI competence to assist project management team in the preparation of construction schedules or perhaps prepare drafts of the construction schedules for them to review and fine tune.

Artificial intelligence (AI) is poised to rapidly transform businesses particularly the construction industry. Although, AI is still a new technology in the construction industry, it has the potential to have a major impact particularly in construction schedules. AI powered scheduling tools could help the project management teams create more accurate and efficient schedules, which could lead to significant cost savings and time savings. Optimized schedules are expected to yield significant cost savings over the actual schedules employed (Kettunen & Kwak, 2018).

Artificial Intelligence has many branches and sub-branches as shown in Figure 1. Artificial Intelligence is the capability of a device to perform functions that are normally associated with human intelligence, such as reasoning and optimization through experience (Grewal, 2014). Artificial intelligence brings into being machines that respond to stimulation consistent with traditional responses from humans, given the human capacity for contemplation, judgment and intention (Grewal, 2014).

A subset of Artificial Intelligence (AI) is Machine Learning (ML) in which intelligence is provided to a system so that it can act automatically make decisions depending on the past experiences (Tiwari, 2022). Machine learning focuses on the development of algorithms that can learn from data without being explicitly programmed. ML algorithms are typically trained on large datasets of labeled data, and they can then be used to make predictions or decisions on new data. One of the types of machine learning is unsupervised learning in which the algorithm is not given any labeled data. Instead, the algorithm is given unlabeled data and it must find patterns in the data on its own. Unsupervised learning algorithms try to infer a function to find hidden relations between data points (Tiwari, 2022).

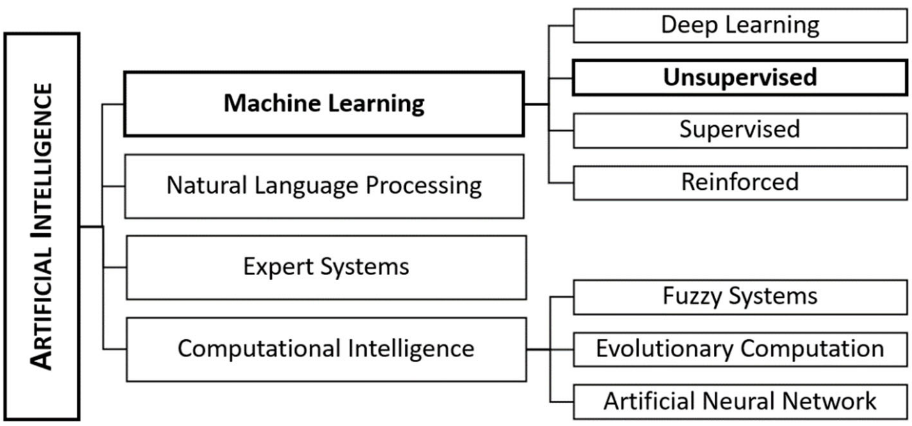

Fig. 1: Sample Areas of Artificial Intelligence

2. RESEARCH METHODOLOGY

A mixed research method was used in this research. The mixed research method draws largely on quantitative and qualitative research (Leedy et al., 2019). Despite its advantages in comparison to mono methods, mixed methods research had been underutilized in the management sciences (Molina-Azorin & Cameron, 2010). However today, mixed methods research is increasingly being used in many disciplines (Bentahar & Cameron, 2015). The use of mixed research method has increased so much that a specialized journal is devoted specifically to mixed methods research - The Journal of Mixed Methods Research, published by Sage (Bentahar & Cameron, 2015). The mixed method was used in this research because both non-numerical and numerical data were needed to evaluate the ability of an AI implementation to comprehend activities of construction schedules. The implementation of the mixed research method was done in four stages: data collection, AI training and preparation, activity interpretation, and analysis as shown in Figure 2.

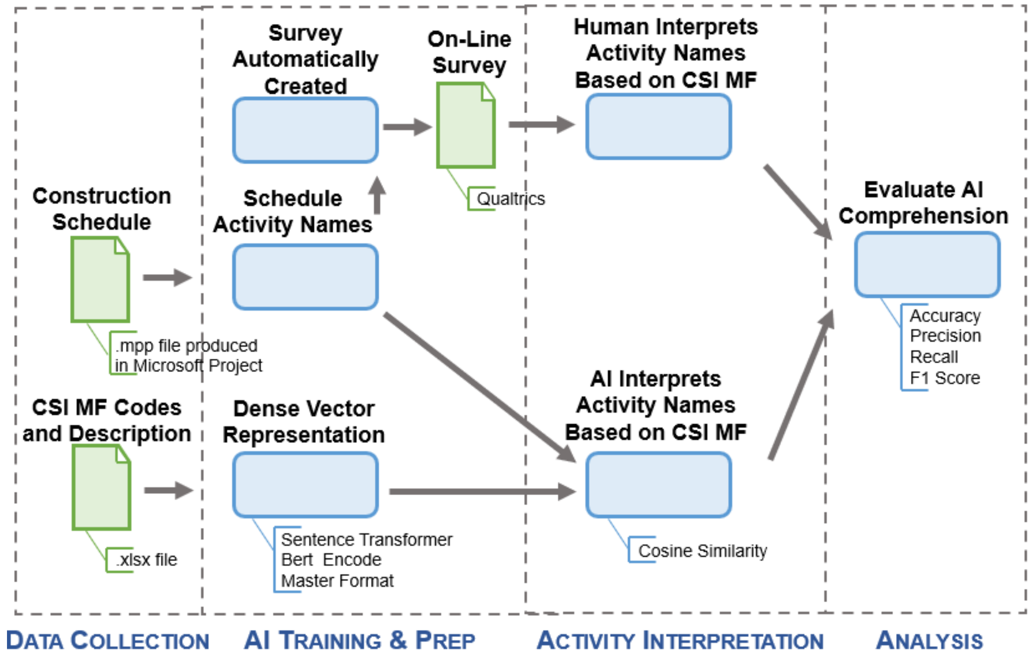

Figure 2: Research Stages

- *Data Collection:* data of a construction schedule as well as data regarding the Construction Specifications Institute (CSI) Master Format (MF) 50 divisions activity codes and descriptions were gathered.

- *Artificial Intelligence Training and Preparation:* the CSI MF 50 divisions activity codes and descriptions were used to train the machine using sentence transformer with a BERT encoder. During the stage also the activities from the construction schedule were extracted and used to automatically create the question of a survey to be deployed on-line through Qualtrics.

- *Activity Interpretation:* the activities of the schedule were provided to a group of humans and an AI. They both were asked to interpret the activities in accordance to the CSI MF 50 divisions. The humans completed the tasks through the online-survey in Qualtrics while the AI completed using the cosine similarity metric.

- *Analysis:* confusion matrix was used to evaluate the AI comprehension of activities in the construction schedule including four metrics – accuracy, precision, recall, and F1 scores to provide a complete picture of the AI performance.

3. RESULTS

3.1 Data Collection

The construction schedule gathered for this research project was composed of 94 activities from notice to proceed to final completion. The project was a 6,500 SF, single story, steel frame, metal stud, gypsum partitions with loadbearing brick and block. The project was a court house in a city in the United States with an approximate cost of $300/SF and a total estimated cost of approximately 2 million dollars.

The Construction Specifications Institute (CSI) Master Format (MF) 50 divisions activity codes and descriptions gathered for this project were composed 7533 individual activities grouped in 35 divisions currently activity from Division 00 – Procurement and Contracting Requirements to Division 48 – Electrical Power Generation.

3.2 Artificial Intelligence Training and Preparation

The training of Artificial Intelligence (AI) was done in Jupyter Notebook which is a free, open-source, interactive web tool known as a computational notebook (Perkel, 2018). Jupyter Notebook was used because it has emerged as a de facto standard for data scientists (Perkel, 2018). The programming code in Jupyter Notebook was done using Python taking advantage of the Sentence Transformers framework to compute semantic similarity and develop the embedding model (Devika et al., 2021). An embedding model is a type of machine learning model that is used to represent words or other discrete entities as real-valued vectors. The real-valued vectors were created using the Bidirectional Encoder Representation Transformers (BERT) Natural Language Inference (NLI) which maps sentences & paragraphs to a 768-dimensional dense vector space and can be used for tasks like clustering or semantic search (Devika et al., 2021). The BERT-NLI models was provided the list of the Construction Specifications Institute (CSI) Master Format (MF) 50 divisions activity codes and returned the corresponding vector for each of the 50 divisions.

The preparation of the survey was done by uploading the construction schedule into a Jupyter Notebook. The Jupyter Notebook extracted the 94 activities from the construction schedule and automatically prepared the 94 questions using the template shown in the Table 1a. Additionally, the questions were grouped into four quartiles according to the cosine of similarity values between the activity and the CSI MF 50 Divisions per the AI Interpretation of the Activities as shown in Table 1b. The four group of questions were uploaded into Qualtrics. In Qualtrics, randomized sub-set of questions to be shown to each participant from each quartile were entered as shown Table 2.

Table 1: Template, Quartiles, and Number of Questions

a. Questions Template
Activity Description: <Activity from Construction Schedule Here> Select from the pulldown below the CSI Master Format Division for which the activity description above (in bold) belongs to. If the activity does NOT belong to any of the CSI Master Format select "None of the Above".

b. Quartile and Number of Questions

Quartiles	**Cos Similarity Values**	**Number of Activities**
Top 25%	More than 0.875	24
Second 25%	0.875 to 0.831	23
Third 25%	0.830 to 0.779	23
Bottom 25	Less than 0.779	24

3.3 Schedule Activity Interpretation

The first part of the construction schedule activity interpretation was done by humans. To ensure that the participating humans could answer the questions within 15 minutes, only the randomized sub-set of questions were provided to each participating human. The sub-sets were composed of 18 of the 94 questions. In the 18 questions, there were three questions from the top two quartiles and six questions from the bottom two quartiles as shown in Table 2. This decision of having more questions from the bottom two quartiles was done because it was anticipated that there was going to be a lower percentage of AI construction activity interpretation that were going to match the interpretation from the participants. Additionally, none of the questions were mandatory, so the participants could skip some of the questions resulting in a total of 316 answers from the participants. The second part of the construction schedule activity interpretation was done by AI using the BERT-NLI model. The AI was given the same construction schedule activities with the same questions given to the human.

Table 2: Questions Template and Number

Quartiles	Questions Per participant	Total Questions Answered
Top 25%	3	51
Second 25%	3	53
Third 25%	6	105
Bottom 25%	6	107
Total	18	316

3.4 Artificial Intelligence Schedule Activity Analysis

The task of interpreting the activities of a construction schedule according to the 50 Divisions of the CSI Master Format was completed first by eighteen human participants. The participants' demographic was as follows: 77.8% Hispanics, 61.1% between 20 and 24 years old, 77.8% males, and 55.6% with 1 to 5 years work experience. The responses of the participants were grouped in the same four quartiles of questions as shown in Table 1 then the answers of the AI were also grouped in according to the four quartiles. If the answer of the AI matched the answer of the humans, the answer was considered a match if not it was considered a no match. The AI identification of the activities match the human answer on average 50% for the first the three quartiles which correspond to the quartiles that the AI was expected to match the human answer. Likewise, the AI identified activities did not match the human answer in the bottom quartile 75% of the times as expected.

Table 3. Questions Template and Number

<table>
<tr><th>Quartiles</th><th>Number of Activities</th><th>Number and % of Match</th><th>Number and % of No match</th><th>Number and % of Match</th><th>Number and % of No match</th></tr>
<tr><td>Top 25%</td><td>24</td><td>14 (58.3%)</td><td>10 (41.7%)</td><td rowspan="3">35 (50.0%)</td><td rowspan="3">35 (50.0%)</td></tr>
<tr><td>Second 25%</td><td>23</td><td>10 (43.5%)</td><td>13 (56.5%)</td></tr>
<tr><td>Third 25%</td><td>23</td><td>11 (47.8%)</td><td>12 (52.2%)</td></tr>
<tr><td>Bottom 25%</td><td>24</td><td>6 (25.0%)</td><td>18 (75.0%)</td><td>6 (25.0%)</td><td>18 (75.0%)</td></tr>
</table>

Furthermore, the Artificial Intelligence (AI) comprehension of the scheduling activities was also done using a confusion matrix. A confusion matrix represents the prediction summary in matrix form (Tiwari, 2022). It is a tool to determine the performance of the AI useful to identify areas where the AI may need improvement. The confusion matrix is useful because shows how many predictions are correct (true) and incorrect (false) per class (Tiwari, 2022). The two classes used in this research were that the AI was either expected to identify (top three quartile) or no identify (bottom quartile) the activities in the construction schedule.

The values used in the confusion matrix for the AI Activity interpretation correspond to the first top three quartiles for identified and the bottom quartile for the not identified. Also, for the actual activity identified corresponds to the match while the not identified correspond to the no match. As shown in Figure 3, the confusion matrix has two rows and two columns with four possible outcomes (true positive, false negative, false positive, and true negative). The top left quadrant shows the number of true positives, which are cases where the AI implementation correctly identified the activity. The bottom left quadrant shows the number of false negatives (also known as type II error), which are cases where the AI implementation was not expected to identify the activity but was in fact able to identify the activity. The top right quadrant shows the number of false positives (also known as type I error), which are cases where the AI implementation was expected to identify the activity, but provided the wrong activity interpretation. The bottom right quadrant shows the number of true negatives, which are cases where the AI implementation was not expected to identify the activity and in fact was not able to identify the activity.

		Actual Activity	
		Identified	**Not Identified**
AI Activity Interpretation	**Identified**	True [1] Positive (TP) 35	False [2] Positive (FP) 35
	Not Identified	False [3] Negative (FN) 6	True [1] Negatives (TN) 18

Legend:

AI was expected to identify activities

TP = True Positive = AI correctly identified activity (as expected)

FP = False Positive = AI did not identify activity

AI was not expected to identify activities

FN = False Negative = AI correctly identify activity (although it was not expected to identify the activity)

TN = True Negatives = AI did not identify activity (as expected)

(1) Correct predictions

(2) Type I Error

(3) Type II Error

Figure 3: AI Interpretation of Construction Activities Confusion Matrix

The confusion matrix information was used to calculate four metrics – accuracy, precision, recall, and F1 scores to provide a complete picture of the AI performance in comprehending the activities in the construction schedule.

- *Accuracy:* is used to find the portion of correctly interpreted activities. In other words, measures how often the AI is correct. The value ranges from 1 for 100% accurate to 0 for 0% accurate. The equation used to calculate accuracy is presented in Equation 1 resulting in the AI having an accuracy to identify activities of 0.56.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{35+18}{35+18+35+6} = 0.56 \qquad \text{Equation 1}$$

- *Precision:* is used to calculate the AI's ability to interpret positive values correctly (True). In other words, measures how often the AI correctly identified the activity when it was expected to do so. Precision is equal to the ratio of the number of construction activities correctly interpreted to the total number of construction activities predicted. The value ranges from 1 for 100% precise to 0 for 0% precision. The equation used to calculate precision is presented in Equation 2 resulting in the AI having a precision to identify activities of 0.50. This result is consistent with literature as fully AI automated approach is still immature to be used in the industry where the best model scored 0.511 precision (Amer et al., 2021)

$$Precision = \frac{TP}{TP+FP} = \frac{35}{35+35} = 0.50 \qquad \text{Equation 2}$$

- *Recall:* (also called sensitivity) is used to calculate the AI's ability to interpret activities among all the activities. In other words, measures how often do the AI correctly identified the activities weather or not is expected to do so. Recall is the ratio of the number of construction activities correctly interpreted to the total number of construction activities interpreted. The value ranges from 1 for 100% recall to 0 for 0% recall. The equation used to calculate precision is presented in Equation 3 resulting in the AI having a recall to identify activities of 0.85.

$$Recall = \frac{TP}{TP+FN} = \frac{35}{35+6} = 0.85 \qquad \text{Equation 3}$$

-*F1-Score:* is the harmonic mean of Recall and Precision. In other words, it is useful when a balance between Precision and Recall needs to be taken into account.

$$F1\ Score = \frac{2*Precision*Recall}{Precision+Recall} = \frac{2*0.50*0.85}{0.50+0.85} = 0.63 \qquad \text{Equation 4}$$

4. SUMMARY

The construction schedule activity names are of particular interest as they are the critical core unit to create the schedule. Unfortunately, there is no standard naming conversion for those activities and they vary from project to project as well as from project management team to project management team. This inconsistency of the activity name makes it extremely challenging for both humans and machines to understand the meaning and scope of the activities. Therefore, the objective of this paper was to evaluate the ability of an Artificial Intelligence (AI) implementation to comprehend activities in a construction schedule. Following a mixed method in this research, the AI was implemented using the Bidirectional Encoder Representation Transformers (BERT) Natural Language Inference (NLI) with the list of the Construction Specifications Institute (CSI) Master Format (MF) 50 divisions activity codes. The result was that the AI has 0.56 accuracy, 0.50 precision, 0.85 recall and, 0.64 F1 Score.

5. FUTURE WORK

Despite the AI not being 100% accurate, this paper opens a wide variety of future research opportunities grounded on the mixed method used in this research with the four stages (Data Collection, AI Training and Preparation, Activity Interpretation, and Analysis). Some of those future research opportunities include: 1- Used other method to evaluate the NLP such as the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve, 2- Evaluating other NLP Encoder, 3- Comparing Performance among multiple NLP Encoders, 4- Implement a mixture of unsupervised and supervise NLP, and 5- Expand the number and type of schedule activities to be implemented just to mention a few. Upon achieving a higher level of comprehension future research could be directed towards using AI to assist humans in the preparation of construction schedules or perhaps prepare drafts of the construction schedules for the human to review.

REFERENCES

Amer, F., & Golparvar-Fard, M. (2019). Automatic Understanding of Construction Schedules: Part-of-Activity Tagging. Proceedings of the 2019 European Conference on Computing in Construction, 1, 190–197. https://doi.org/10.35490/EC3.2019.196

Amer, F., Jung, Y., & Golparvar-Fard, M. (2021). Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. Automation in Construction, 132, 103929. https://doi.org/10.1016/j.autcon.2021.103929

Bentahar, O., & Cameron, R. (2015). Design and Implementation of a Mixed Method Research Study in Project Management. The Electronic Journal of Business Research Methods. https://www.semanticscholar.org/paper/Design-and-Implementation-of-a-Mixed-Method-Study-Bentahar-Cameron/7541f31e4c91c95e68a83f89fa15edb7eba30b63

Devika, R., Vairavasundaram, S., Mahenthar, C. S. J., Varadarajan, V., & Kotecha, K. (2021). A Deep Learning Model Based on BERT and Sentence Transformer for Semantic Keyphrase Extraction on Big Social Data. IEEE Access, 9, 165252–165261. https://doi.org/10.1109/ACCESS.2021.3133651

Essam, N., Khodeir, L., & Fathy, F. (2023). Approaches for BIM-based multi-objective optimization in construction scheduling. Ain Shams Engineering Journal, 14(6), 102114. https://doi.org/10.1016/j.asej.2023.102114

Grewal, D. S. (2014). A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering. IOSR Journal of Computer Engineering, 16, 09–13.

Halpin, D. W. & Senior, Bolivar. (2017). Construction management (5th Edition). John Wiley & Sons.

Heigermoser, D., García de Soto, B., Abbott, E. L. S., & Chua, D. K. H. (2019). BIM-based Last Planner System tool for improving construction project management. Automation in Construction, 104, 246–254. https://doi.org/10.1016/j.autcon.2019.03.019

Hong, Y., Xie, H., Bhumbra, G., & Brilakis, I. (2021). Comparing Natural Language Processing Methods to Cluster Construction Schedules. Journal of Construction Engineering and Management, 147(10), 04021136. https://doi.org/10.1061/(ASCE)CO.1943-7862.0002165

Hong, Y., Xie, H., Hovhannisyan, V., & Brilakis, I. (2022). A graph-based approach for unpacking construction sequence analysis to evaluate schedules. Advanced Engineering Informatics, 52, 101625. https://doi.org/10.1016/j.aei.2022.101625

Kettunen, J., & Kwak, Y. H. (2018). Scheduling Public Requests for Proposals: Models and Insights. Production and Operations Management, 27(7), 1271–1290. https://doi.org/10.1111/poms.12868

Leedy, P. D., Ormrod, J. E., & Johnson, L. R. (2019). Practical research: Planning and design (Twelfth edition). Pearson.

Magalhães-Mendes, J. (2011). A two-level genetic algorithm for the multi-mode resource-constrained project scheduling problem. 5, 271–278.

Molina-Azorin, J., & Cameron, R. (2010). The Application of Mixed Methods in Organisational Research: A Literature Review. Electronic Journal of Business Research Methods, 8, 95–105.

NYDOT. (2016). Best Practices for CPM Schedule Specification Compliance (p. 53). New York Department of Transportation. https://www.dot.ny.gov/main/business-center/contractors/construction-division/construction-repository/Best_Practices_for_CPM_Schedule_Spec_Compliance.pdf

Okonkwo, C., Garza, R., Sulbaran, T., & Awolusi, I. (2022). A Review of Genetic Algorithm as a Decision-Making Optimization Tool in Project Management. EPiC Series in Built Environment, 3, 254–262. https://doi.org/10.29007/jzcl

Perkel, J. M. (2018). Why Jupyter is data scientists' computational notebook of choice. Nature, 563(7732), 145–147. https://doi.org/10.1038/d41586-018-07196-1

Rosłon, J., Książek-Nowak, M., & Nowak, P. (2020). Schedules Optimization with the Use of Value Engineering and NPV Maximization. Sustainability, 12(18), 7454. https://doi.org/10.3390/su12187454

Sulbaran, T., & Ahmed, F. (2017). Expert System for Construction Scheduling Decision Support Based on Travelling Salesman Problem. 53rd ASC Annual International Conference Proceedings. http://ascpro.ascweb.org/chair/paper/CPRT223002017.pdf

Tiwari, A. (2022). From theory to applications—Chapter 2—Supervised learning: In R. Pandey, S. K. Khatri, N. kumar Singh, & P. Verma (Eds.), Artificial Intelligence and Machine Learning for EDGE Computing (pp. 23–32). Academic Press. https://doi.org/10.1016/B978-0-12-824054-0.00026-5

Yurchyshyna, A., & Zarli, A. (2009). An ontology-based approach for formalisation and semantic organisation of conformance requirements in construction. Automation in Construction, 18(8), 1084–1098. https://doi.org/10.1016/j.autcon.2009.07.008