Article

Investigating the Use of ChatGPT for the Scheduling of Construction Projects

Samuel A. Prieto * , **Eyob T. Mengiste and Borja García de Soto**

S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, Abu Dhabi P.O. Box 129188, United Arab Emirates; eyob.mengiste@nyu.edu (E.T.M.); garcia.de.soto@nyu.edu (B.G.d.S.)
* Correspondence: samuel.prieto@nyu.edu

Abstract: Generative Pre-Trained Transformer (GPT) language models such as ChatGPT have the potential to revolutionize the construction industry by automating repetitive and time-consuming tasks. This paper presents a study in which ChatGPT was used to generate a construction schedule for a simple construction project. The output from ChatGPT was evaluated by a pool of participants that provided feedback regarding their overall interaction experience and the quality of the output. The results show that ChatGPT can generate a coherent schedule that follows a logical approach to fulfill the requirements of the scope indicated. The participants had an overall positive interaction experience and indicated the potential of such a tool in automating many preliminary and time-consuming tasks. However, the technology still has limitations, and further development is needed before it can be widely adopted in the industry. Overall, this study highlights the advantages of using large language models and Natural Language Processing (NLP) techniques in the construction industry and the need for further research.

Keywords: natural language processing; ChatGPT; scheduling; generative pre-training transformer; project management; Construction 5.0; GPT 3.5

1. Introduction

Natural Language Processing (NLP) combines areas such as linguistics, computer science, and Artificial Intelligence (AI) and focuses on the interaction between computers and humans using programs that are developed from large natural language data [1]. Selected applications of NLP in the construction industry include extracting information from construction documents such as specifications, plans, and contracts and converting it into a structured format that can be quickly processed by computers [2]. This can help streamline the process of reviewing and comparing construction documents and reduce the risk of errors caused by manual data entry. NLP has also been used to analyze construction site data such as progress reports, safety inspections, quality control reports, and code compliance checking and extract insights that can help improve project efficiency and mitigate risks [3–8]. NLP-powered chatbots and virtual assistants can be used to improve communication on construction sites by providing a quick and convenient way for project stakeholders to access information and ask questions about the project [9].

In the context of project scheduling, plenty of tools are available in the market for optimizing and facilitating the management of projects (e.g., Microsoft Project, Primavera P6, ALICE, and SYNCHRO, to name a few). In general, the information is schematized and structured in charts, tables, and diagrams. Few studies have investigated the application of natural language to develop project schedules. One example includes the work by Amer et al. [10] that used an NLP-based approach to automate the connection between look-ahead and master-schedule activities. Their method was used by project participants to create look-ahead plans from the description of activities in the master schedule. It could

be expected that, with enough information about the project being provided to the system, NLP techniques could be used to generate or enhance construction schedules based on project details (i.e., the scope of work) provided by a user.

In summary, NLP has the potential to significantly improve efficiency, accuracy, and overall communication in the construction industry. Within the field of NLP, research has been focused on developing Large Language representation Models (LLM) that expand the applicability of NLP to more detailed human language understanding tasks such as translation, text classification, and holding conversations [11,12]. ChatGPT is based on a Generative Pre-Trained Transformer model, one of the multiple available LLMs within the NLP research field.

Aims and Contribution

In this paper, we evaluate the applicability of a Generative Pre-Training Transformer language representation model (GPT) to assist in developing an automated construction schedule based on natural language prompts. The aims of this study are summarized as follows:

- Explore the possible applications and limitations of a rapidly growing and powerful tool for construction scheduling and resource loading.
- Conduct a preliminary case study involving multiple users applying GPT to generate a resource-loaded project schedule for a simple project based on a given detailed natural language description input (i.e., prompt).
- Evaluate the results obtained from the participants in the case study based on parameters such as accuracy, efficiency, clarity, coherence, reliability, relevance, consistency, scalability, and adaptability.

To the authors' knowledge, this is the first study that uses a GPT model for construction scheduling and the first one to explore the different applications of a conversation-focused LLM, such as ChatGPT, in the construction field.

The rest of the paper is organized as follows: Section 2 provides a brief state-of-the-art study on various language representation models and the current trend in project management tools. Section 3 presents the methodology followed to evaluate the results from the case study. Section 4 presents the case study. Section 5 discusses the results and findings from the case study, going into detail about the strengths and limitations of the tool. Finally, Section 6 contains some conclusions and future work by the authors on the application.

2. Literature Review

2.1. Language Representation Models

Large language models have gained significant attention in recent years due to their ability to generate human-like text and perform a wide range of language-based tasks. In the construction industry, large language models have the potential to improve efficiency, accuracy, and communication in several different ways. The fields where NLP technology has been tested and applied are limited. The application of NLP technologies in the construction sector is limited. Some examples include the work by Xue et al. [2] that used NLP techniques to summarize construction contracts. Locatelli et al. [13] studied the potential of combining NLP and BIM, focusing on automated compliance checking and semantic BIM enrichment goals.

Regarding language representation models, different approaches have been developed in the past decade [12]. The major ones can be grouped into three families: the autoregressive language model Generative Pre-trained Transformer (GPT), the Bidirectional Encoder Representations from Transformers (BERT) and the Multi-Task Learning.

2.1.1. Generative Pre-Trained Transformer (GPT)

The GPT family (Generative Pre-trained Transformer) of language models was developed by OpenAI. The models are trained on a large dataset of text and, as expected, are able

to generate human-like text. At the time of this article, the most recent version of the GPT family is GPT-3 [14], which is the base for ChatGPT (considered as GPT-3.5), the tool used for this study. GPT-3, released in June 2020, has 175 billion parameters, making it one of the largest language models to date. Since its release, it has been proven that GPT-3 is capable of performing some tasks that traditionally require human-level understanding, such as writing essays and programming. Despite the technology not being particularly new [15], GPT-3.5 has been fine-tuned for information retention during the conversation, making it suitable for the scope being tested in this study. This feature has motivated researchers to study the possibility of incorporating such language models into activities that were solely reserved for human–human interaction, such as healthcare delivery [16]. Floridi and Chiriatti [17] studied the scope, limits, and consequences of the GPT-3 model and how society will have to get used to not being able to tell if a text was written by a human or AI.

2.1.2. Bidirectional Encoder Representations from Transformers (BERT)

BERT (Bidirectional Encoder Representations from Transformers) is a pre-trained transformer-based neural network model developed for NLP tasks, such as text classification and question answering (chatbots). BERT was first introduced by Google in 2018 [18] and has been the state-of-the-art for many of the NLP models developed afterward. Hassan et al. [19] used a BERT-Based model to identify risky and hazardous situations related to construction. Moon et al. [20] used a BERT model to automatically detect contractual risk clauses from construction specifications. Their approach classified contractual risk categories to provide reviewers with crucial clauses that commonly cause disputes, such as payment, temporal, procedure, safety, role and responsibility, definition, and reference. Yao and Garcia de Soto [21] used a BERT model classification for semantic screening trained on construction-specific documentation to investigate the main topics related to construction cybersecurity.

2.1.3. Text-to-Text Transfer Transformer (T5) Multi-Tasking Learning

T5 (Text-to-Text Transfer Transformer) is a multi-task pre-training model for NLP tasks. T5 was introduced by Google in 2020 [22]. T5 models are trained to perform multiple tasks at once, which allows them to learn a general-purpose understanding of language that can be fine-tuned for specific tasks.

2.2. Automation of Construction Scheduling

The area of schedule automation has been heavily researched in the past few decades. Modern construction project scheduling can generally be categorized as BIM-Driven and Machine Learning-based schedule generations.

2.2.1. BIM-Driven Schedule Generation

BIM models comprise geometric information (special relationships of components), materials, and resources. Under the current industrial practice, project management teams utilize the information from the BIM to optimize and generate schedules. Researchers improved the process of manual schedule development by automating task dependency and sequencing using pre-set rules, patterns, or pre-set knowledge learned from historic cases [23,24].

These methods extract required data inputs from the pre-built information model to produce a schedule with a logical order [25]. However, BIM models often do not include environmental factors, temporary structures, equipment and material availability, specific construction specifications, and the methodology of the building site. Therefore, the BIM-based automatic generation of schedules requires manual intervention to be practical [25,26].

2.2.2. Machine Learning-Based Schedule Generation

Several approaches have been proposed for performing predictive modeling to utilize historical data to predict project outcomes [27]. However, these approaches generally rely on large amounts of data to train the models. The source data could be visual [28], where the algorithm understands the characteristics of the sites and work performance.

Language-based methods were also the focus of research for automated construction schedule development. Natural language models such as GPT [10] or language clustering methods such as Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA) [29] were used to develop task dependency and clustering from information derived using natural human language.

Hong et al. [26] proposed a graph-based construction scheduling approach. Their proposed approach stores past best practices and recycles the information to optimize the resource usage, task sequence, and duration of the new project.

In general, Machine Leaning-based approaches produced promising results for construction schedules; however, the performance of most methods depends on the availability of data and the data processing technical and infrastructural capacity of the user.

3. Methodology

A small experiment has been designed to evaluate the potential of ChatGPT as an aid in project management in construction, focusing on project scheduling and task assignment. The experiment consists of a simple construction project (scope of work) in an existing space. The goal is to retrieve a logical and accurate task breakdown from ChatGPT. The same experiment was performed by different participants, allowing them to challenge the tool and modify the original plan to evaluate the ChatGPT response. The profile of the different participants is summarized in Table 1.

Table 1. Summary of the profile of the participants in the experiment.

Participant	Experience	Academic Background
1	2+ years in the academic field	Civil Engineering, M.Sc.
2	2+ years in the industry 2+ years in the academic field	Civil Engineering, M.Sc.
3	7+ years in the academic field 2+ years in the industry	Civil Engineering, Ph.D.
4	10+ years in the academic field 15+ years in the industry	Civil Engineering and Construction Management, Ph.D.
5	2+ years in the academic field	Civil Engineering, B.Sc.
6	7+ years in the academic field 2+ years in the industry	Electrical Engineering, Ph.D.

Various parameters were used to evaluate the results of these experiments, including accuracy, efficiency, clarity, coherence, reliability, relevance, consistency, scalability, and adaptability. Accuracy was measured by comparing the ChatGPT-generated project schedules and task assignments to a baseline schedule and assignments created by a human project manager. Efficiency was evaluated by collecting data about the time required to create a schedule and assign tasks using ChatGPT, as well as the time needed to fix the number of errors or mistakes made throughout the process. Clarity and relevance were assessed by evaluating the proposed schedules and responses subject to modifications and complexity changes to ensure they were clear and easy to understand. The responses must be relevant to the input prompt and not deviate from the original task. Coherence was measured by examining the results, such as task dependencies or crew assignments, to ensure that they made sense from a logical point of view in the eye of an expert. Reliability was evaluated by checking conformity with standards and assessing its adequateness if

those results were to be used in a real project. Consistency was measured by evaluating the invariance of the ChatGPT responses based on slight modifications to the initial prompt or multiple instances of the same prompt. The scalability and adaptability of the ChatGPT solution were also evaluated, including its ability to handle larger projects or deal with additional tasks or responsibilities, as needed, and its ability to handle changing project requirements or unforeseen challenges.

The initial input for the study consisted of a paragraph with enough details of the scope to be completed (see Section 4). Details included the description of the work to be carried out, the dimensions and the type of material to be used, the expected level of completeness of the task, and the due date for the task to be completed. The results expected from the process were a list of tasks and overall scheduling (sequence) of the project, with the ability to interactively react to changes made by the user. An overview of the methodology is shown in Figure 1 (following BPMN notation).
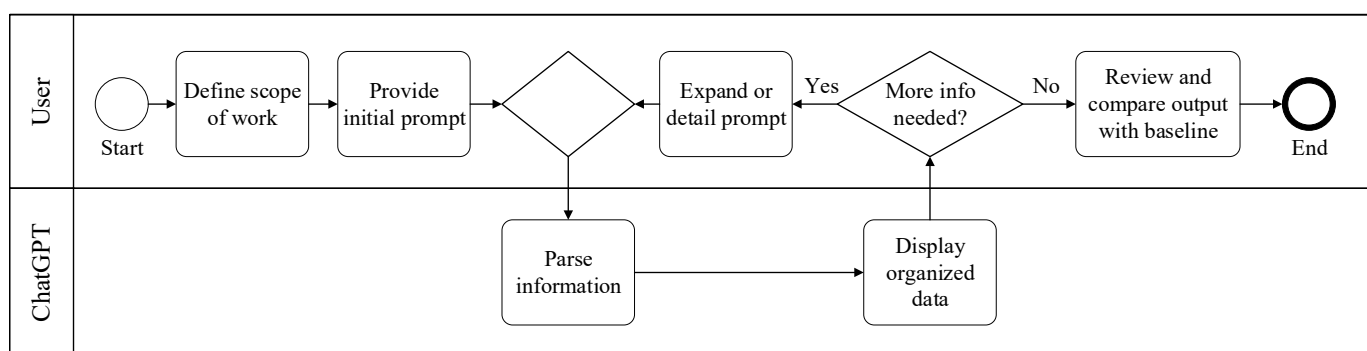
Figure 1. Overview of the main steps for using and assessing ChatGPT in this study.

The original scope underwent small modifications concerning the original prompt to challenge ChatGPT. Some modifications included adding a new scope (e.g., electrical or plumbing work).

A survey was conducted containing instructions to set the bases of the experiment and a series of questions to evaluate the quality of the output generated by ChatGPT and the participants' experience interacting with it. The survey took place from 9 to 14 January 2023. Participants' backgrounds and experiences ranged from civil engineering and construction management to industrial engineering. The results are summarized in Section 5.

4. Case Study

A simple project and scope were used to evaluate the performance of ChatGPT when providing a construction schedule. The ChatGPT versions used were the Dec 15 and Jan 9 Versions. In general, the work consisted of the addition of a partition wall in an existing space. A simplified floorplan with the key elements of the scope of work is shown in Figure 2a. To ensure that all the participants provided the same information regarding the scope, an initial prompt was given to ChatGPT in all cases. The initial prompt was:

"A set of instructions on a construction project will be provided. You will store the provided information, and you won't provide any answers to the initial prompt until asked otherwise".

The initial input describing the scope of work was:

"A new partition needs to be done in an already existent space, where the new partition is grouted with the existing walls. The details of the room to be partitioned are the following: the room is rectangular shaped, 4 m by 4 m in total. The walls are made of concrete masonry units. The height of the walls is 3 m, and the width is 20 cm. The new partition needs to be made out of concrete masonry units as well. The partition is meant to split the original space in half, resulting in two individual spaces of approximately 4 m by 2 m.

The partition needs to account for the installation of a single solid, two-panel wooden door of 0.8 m in width by 2.1 m in height and 35 mm thickness that will communicate the two new spaces. After the partition is made, it needs to be plastered with two layers of stucco and painted with two layers of white latex paint on both sides of the wall. No electrical or plumbing installation is needed. No ceiling work is needed. The floor is cement screed. The work needs to be completed in less than three weeks".

(a)

(b)
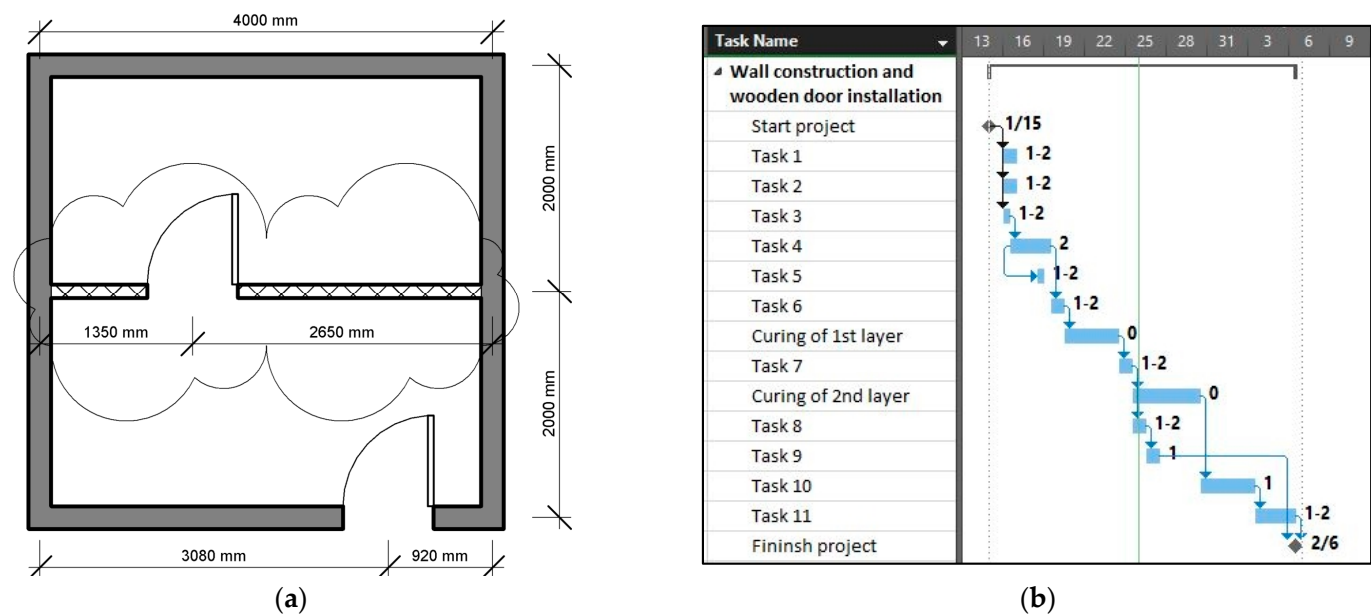
Figure 2. (**a**) General view of the floorplan and (**b**) Gantt chart of main activities used as the baseline for the required scope summarized in Table 2.

A typical list of activities for the proposed scope is shown in Table 2. The logic of the tasks is represented in the Gantt Chart shown in Figure 2b. This was used as the baseline for comparison with the output generated by ChatGPT. Based on the initial prompt, ChatGPT was asked to use that information to generate a construction schedule to complete the scope. To do that, the following prompt was asked to ChatGPT:

"Can you come out with a suitable project schedule"?

To structure the obtained information into usable data, the following structure was asked from ChatGPT:

"Based on the details of the work to be completed, extract the information in the following structure 'task name/task priority/task dependencies/number of people needed/expected duration of task."

The prompt *"Add this information into columns"* could be used to show the output in a table format. Based on this simple scope, a series of questions with a set of initial instructions were put in a survey form. This form was distributed to several individuals (participants) with different skill sets and qualifications working in the construction and AI fields. The feedback from the six participants is summarized in Section 5. Examples of the interface of these prompts in ChatGPT are shown in Appendix B (Figure A1).

5. Results and Discussion

The entire output regarding the schedule obtained by the six different participants is summarized in Appendix A (Table A1). The table contains all the information regarding the different tasks proposed by ChatGPT, their dependencies, the assigned priority, the expected number of people, and the time needed to complete them. In general, ChatGPT provided a logical (although very linear) sequence of tasks in all cases. The tool could extrapolate a breakdown of the steps needed without that information

being explicitly provided and establish logical and coherent dependences amongst the proposed tasks. At first glance, the output seemed coherent and reasonable, and the participants were awed by the speed at which the responses were provided (in most cases, within a few seconds of entering the prompt). However, with further inspection, it was clear that not all of the proposed tasks agreed with the scope of work. Both quantitative and qualitative analyses have been made to assess the performance and interaction with ChatGPT.

Concerning the quantitative results, Table 3 shows a direct comparison of the tasks proposed by ChatGPT in all six cases with respect to the baseline. It is worth noting that the tasks related to the wooden door (i.e., the frame installation and its protection before painting) were not considered in any of the schedules proposed by ChatGPT. This is due to the fact that ChatGPT has not been trained for specific construction purposes, and it is not aware that, for the door installation, the frame needs to be placed first. In some cases, the two layers of plastering are proposed as one single task, and most of the planning and preparation are joined into a single task.

In addition, three of the six responses by ChatGPT included the demolition of the existing wall, which is not relevant to the scope provided. This was probably wrongfully inferred by ChatGPT based on the information about a *"new partition needs to be done in an already existent space"*, and demolition might have been assumed because of the existing space condition. Additionally, incorrect information was provided regarding the tasks that would be expected. For example, in three of the six cases, ChatGPT indicated the need for foundation work (e.g., "Excavation and preparation of foundation for new partition"; Participants 2, 3, and 4 in Table A1), which might not have been required for a partition wall (non-load bearing). In one case, it also suggested the installation of steel rebar for the new partition and the placement ("pouring") of concrete for the new partition. However, minor schedule errors can be fixed by conversing with ChatGPT, instructing the tool to rearrange some of the tasks, asking for a more detailed breakdown, or providing more information that was not properly understood, such as the frame installation for the door.

The full data regarding the performed survey, with all the responses from the participants and their full conversation with ChatGPT, can be made available to interested readers upon reasonable request to the authors. The proposed dependencies and sequence of tasks are logical for the most part, despite being very linear. The proposed sequence for installing the wooden door is arguably not ideal. ChatGPT includes the door installation right after the wall erection. In general, the sequence for installing the door would be after the plastering, as indicated in the baseline schedule, to avoid damage from intermediate tasks. If only the door frame is considered, the suggested sequence would be adequate; however, a clear breakdown for that was not provided or inferred by ChatGPT.

In order to evaluate both the duration and crew needed for each task, only those tasks present in the baseline were considered. The durations for each task reported by ChatGPT are summarized in Table A1. The variation between the baseline durations (Table 2) and the ones from ChatGPT is displayed in Figure 3. The bars represent the difference between the time proposed by ChatGPT in each of the different survey participants' experiments and the one present in the baseline.

Table 2. List of tasks derived from the scope of work indicated in the original prompt and considered as the baseline for this study.

Task No.	Task Name	Task Dependencies	People Needed *	Expected Duration
1	Inspect the existing space and check proposed work is in line with existing conditions	-	1–2	1 day
2	Prepare the work area and protect surrounding areas as needed	1	1–2	1
3	Measure and mark the location of the new partition, including the location for opening (door)	1	1	1
4	Install CMU for new partition	1	2	3
5	Install framing for the new door	4	1–2	1
6	Apply the first stucco layer to the CMU wall—includes curing time	4 SS + 2	1–2	3
7	Apply the second stucco layer to the CMU wall—includes curing time	6	1–2	4
8	Install and adjust the wooden door	7	1	0.5
9	Protect the door in preparation for the painting of the new CMU partition wall	8	1	0.5
10	Finish wall (prime, paint, apply two layers—allow drying time per manufacturer's recommendations)	7	2	4
11	Clean-up and final inspection	10	1	1
			TOTAL	15.5 days **

* RS Means was considered when estimating the people needed. ** Total duration, including weekends, as per Figure 2b (equivalent to 11.5 workdays).

The bars without data are for tasks that were not considered by ChatGPT. For a given task, a positive increment indicates that the duration from ChatGPT is greater than the one from the baseline, and vice versa. Values equal to 0 indicate that the duration from ChatGPT is the same as the one in the baseline. For example, for Task 4, participant 1 (P1) did not have this task, so the value for that is empty (i.e., bars without data). For P2 and P6, the duration obtained from ChatGPT was the same as the baseline (i.e., the deviation is 0). For P3, the duration from ChatGPT was 1 day more than the baseline (3 days in the baseline vs. 4 days in ChatGPT). For P4 and P5, the duration from ChatGPT was 1 day less than the baseline (3 days in the baseline vs. 2 days in ChatGPT). Some of the tasks with the biggest difference are those that involve drying time (i.e., Tasks 6, 7, and 10). In the baseline, drying time has been considered for both the plastering and painting tasks, which is why those tasks take more than three days each. However, ChatGPT does not seem to consider drying times.

The estimation of the number of workers reported by the participants using ChatGPT was compared with the baseline estimation. Most of the numbers of workers per task from all the ChatGPT responses and the baseline are given as a range of maximum and minimum numbers. Therefore, the comparison was carried out considering the maximum and the minimum possible difference in workers per task per participant. The highest deviation from the baseline estimation is two workers. The results are shown in Figure 4. Positive increments indicate an increase in the number of workers compared to the baseline, and vice versa. For instance, the estimated number of workers for Task 6 in the baseline ranged from one to two. P1 and P4 reported that ChatGPT estimated the same amount. However, for P3, ChatGPT estimated that Task 6 would require two workers. Therefore, based on the difference between the baseline (considering the range) and this response, ChatGPT estimated that Task 6 would require a maximum of one more worker (in the case

of the lowest baseline estimate) and a minimum of zero additional workers (in the case of a maximum estimate), as represented by the bar for P3 in Task 6. Similarly, P5 and P6 estimated that a minimum of two and a maximum of three workers would be required to complete Task 6. This estimate is one worker higher than the baseline estimate, in the case of both upper and lower baseline estimates.

Table 3. Comparison of the tasks proposed in the schedules generated by ChatGPT with the ones from the baseline.

Task No.	Task Name (Baseline)	Participant					
		1	2	3	4	5	6
1	Inspect the existing space and check proposed work is in line with existing conditions	✗	✗	✗	✗	✓	✓
2	Prepare the work area and protect surrounding areas as needed	✗	✗	✓	✗	✓	✓
3	Measure and mark the location of the new partition, including the location for opening (door)	✗	✗	✓	✗	✓	✓
4	Install CMU for new partition	✗	✓	✓	✓	✓	✓
5	Install framing for the new door	✗	✗	✗	✗	✗	✗
6	Apply the first stucco layer to the CMU wall—includes curing time	✓	✓	✓	✓	✓	✓
7	Apply the second stucco layer to the CMU wall—includes curing time	✓	✓	✓	✓	✓	✓
8	Install and adjust the wooden door	✓	✓	✓	✓	✓	✓
9	Protect the door in preparation for the painting of the new CMU partition wall	✗	✗	✗	✗	✗	✗
10	Finish wall (prime, paint, apply two layers—allow drying time per manufacturer's recommendations)	✓	✓	✓	✓	✓	✓
11	Clean-up and final inspection	✓	✓	✓	✓	✓	✓

✓ Included by ChatGPT. ✗ Not included by ChatGPT.
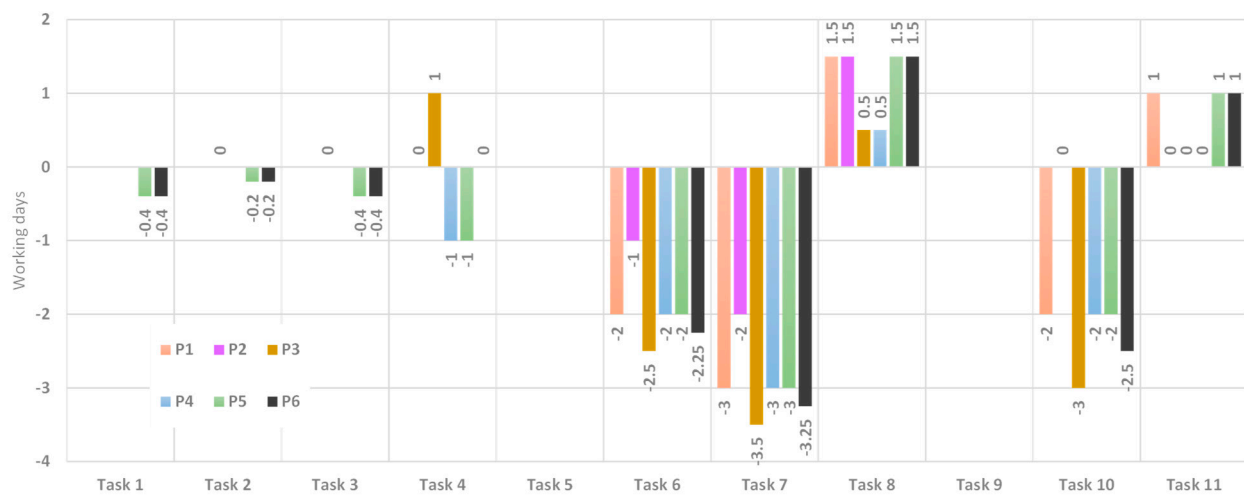
Figure 3. Deviation between the durations proposed by ChatGPT and the baseline for each task from each participant (P).

To test the adaptability and scalability of ChatGPT, the original prompt was modified to add additional components to the scope, such as electrical or plumbing work. For example:

"The new wall that needs to be built for the partition needs to have a couple of electrical sockets. Can you include that task in your schedule and regenerate it"?

ChatGPT was able to modify the original list of tasks and update the sequence accordingly. An example of the new task for the example above is shown in Table 4.

To further challenge ChatGPT, information regarding the cost was also asked for. The following prompt was used:

"Based on the information above, can you provide a cost estimate for labor and materials needed"?

As expected, the responses from ChatGPT were vague and very generic, constantly advising to consult with a licensed contractor to obtain a rough cost estimation. A sample response from ChatGPT was, *"I'm sorry, but I am not able to provide you with a cost estimate for labor and materials needed based on the information provided. The cost of construction projects can vary greatly depending on a variety of factors such as location, materials, labor costs, and the specific requirements of the project. Additionally, factors such as the cost of permits, inspections, and any necessary equipment rentals should also be considered. It would be best to consult with a licensed contractor or estimator to provide you with an accurate cost estimate for your project".*

When insisting on providing a rough estimate, the response was still generic (e.g., *"As a language model AI, I don't have access to current market pricing and cost may vary greatly depending on location, materials, and labor. However, for a rough estimate for a partition wall and door installation in a standard room, you could expect to spend anywhere from $1000 to $5000 or more for a standard partition wall and door installation. This is a rough estimate and the real cost could vary greatly depending on the location, the specific needs of the project, and the contractor you choose. It is important to consult with a local licensed contractor or estimator to provide you with an accurate cost estimate for your project"*). This shows that ChatGPT will provide a response even if the information is not reliable and illustrates one of the big limitations of the current model: the data are not updated since it is not connected to real-time internet data. Additionally, since the model is taught to be aware of its own limitations, the default response from ChatGPT provides very generic and boilerplate statements, such as relying on a licensed contractor for such a query.
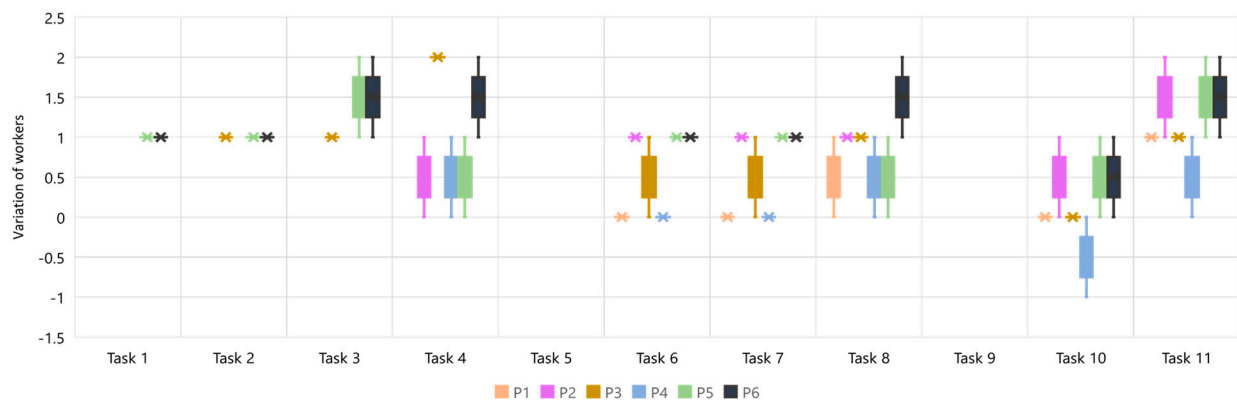
Figure 4. Deviation of the needed workers proposed by ChatGPT with respect to the baseline.

Regarding the quality of the output generated by ChatGPT, the participants' general impression was positive. They were impressed by the fact that ChatGPT could produce a (for the most part) logical and coherent task breakdown with little initial information, despite being a model not trained specifically for construction purposes. The results are summarized in Figure 5.

Table 4. Information regarding the installation of a couple of electrical sockets.

Task Name	Dependencies	People Needed	Expected Duration
Install electrical sockets	Building the new partition	2	4 h

The overall evaluation is very positive, with ratings of "good" and "very good" predominating in all the different features. Accuracy and reliability received the lowest scores due to the issues related to using tasks that were not related to the scope of work (e.g., excavation, foundation work, rebar, etc.) when compared to the baseline, therefore making the results not as reliable as they would need to be in a professional environment.

In addition to the evaluation of the performance of ChatGPT regarding the proposed tasks and sequence (i.e., scheduling component), a set of qualitative questions were asked to evaluate the overall impressions of the participants regarding their experience interacting with ChatGPT. This included their impression of the type of communication over the classical methods (i.e., charts and tables) and their feedback about potential uses for ChatGPT in the construction field. The results regarding the interaction with ChatGPT are presented in Figure 6. Overall, the interaction was very positive, with most participants rating different aspects of the interaction "very good", such as how intuitive it was, how comfortable it was, how efficient it was, and the overall interaction experience.
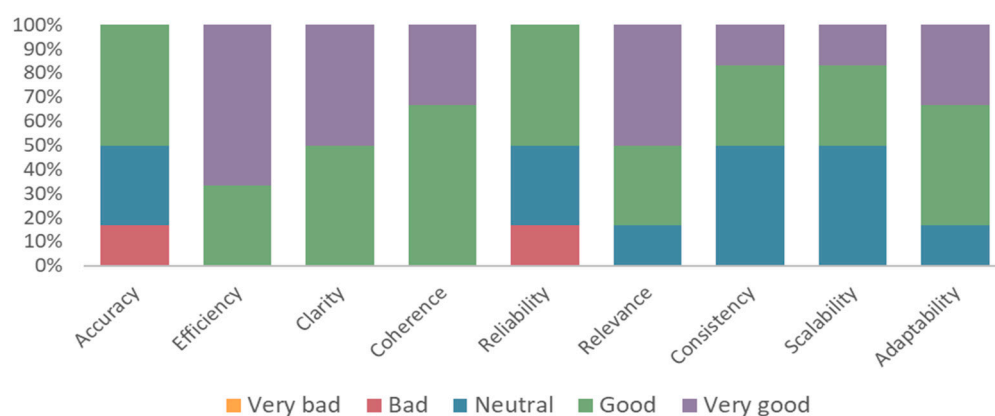
Figure 5. Results from the survey regarding the evaluation of the ChatGPT output.

Regarding the preference for this communication (i.e., in a dialog format) instead of the classical methods (lists, tables, charts), all the surveyed participants shared a common opinion of tables and charts being more intuitive and easier to read than the results provided by ChatGPT. Nonetheless, they all agreed that despite this not being the preferred final form for displaying the results, it might become an important and useful tool for extracting the information to be displayed.
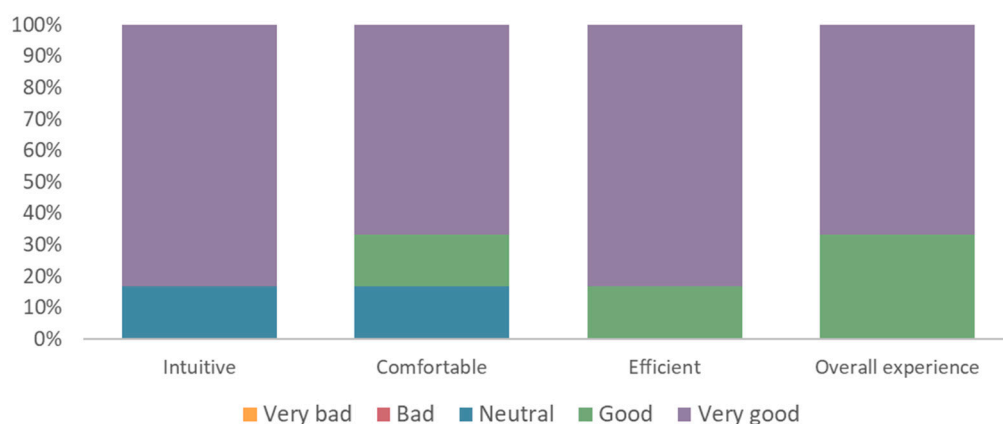
Figure 6. Results from the survey regarding the evaluation of the interaction with ChatGPT.

The possible uses proposed by the surveyed participants ranged from initial consulting, safety and risk identification, basic design, cost estimation, processing and evaluation of contracts, or identifying vulnerabilities.

Overall, they all agreed that any task involving text processing could benefit from the inclusion of Natural Language Processing techniques.

Limitations

The presented case study has been useful for preliminary testing and evaluating the possible uses and capabilities of ChatGPT when applied to the generation of a schedule for simple projects. The results are promising; however, the complexity of the case study is very limited. Further studies considering increased complexity to resemble actual construction projects should be conducted to further consider this technology in the construction field.

The results provided are based on a limited number of participants and should not be used for generalization purposes. A bigger statistical pool of surveyed people is needed for a wider generalization of the results obtained.

6. Conclusions and Future Work

Since ChatGPT has been made available, it has attracted researchers from different fields. Natural Language Processing techniques such as GPT models have been around for some time, but ChatGPT (GPT 3.5) differentiates from them by being specifically trained for conversation, instead of single-prompt usage. That feature allows for a much more fluent interaction and the ability of the user to modify and change the design parameters on the go. The application of such tools in the construction industry should not be overlooked. This study consisted of a simple example to assess the applicability of such a tool in the context of project scheduling. The performance and results were very promising for the simple use case, especially considering that ChatGPT has not been specifically trained for such an application. However, several significant flaws would limit the application of such a tool in a real project. Having said that, the overall performance was reasonable, and the interaction experience was positive. This is important to show that such technology could be relatively easy to integrate if consistent and reliable performance is achieved. Specialized tools in a given field (e.g., project scheduling) could become extremely beneficial and play an important role in the automation of repetitive and time-consuming parts.

Future research needs to be conducted to further explore the applicability and capabilities of Natural Language Processing tools in the construction industry. To do that, a GPT model will be specifically trained for construction purposes and challenged with a more complex scenario. A larger pool of participants will be surveyed in a future study to more accurately generalize the results. A larger and more complex case study is also needed to properly validate the results, which is why this will be tackled in future work. In addition, the way in which the initial information is given to ChatGPT could become a nuisance for larger and more complex projects. This is why it is worth exploring new methods for inputting information into these models, such as in the form of floorplans and images.

Appendix A

Table A1. Summary of tasks and related information from the six participants during their interaction with ChatGPT using the original prompt describing the scope of work.

Participant	Task Name	Dependencies	Priority *	People Needed	Expected Duration
1	1. Demolition of existing wall	-	H	2–3	1 day
	2. Framing of partition	1	H	2–3	2 days
	3. Measurement and cutting of wooden door	2	M	1–2	1 day
	4. Installation of wooden door	3	M	1–2	1 day
	5. Plastering of partition	4	L	1–2	2 days
	6. Final touch-ups and clean-up	5	L	2	2 days
	7. Painting of partition	5	L	2	2 days
				TOTAL	11 days
2	1. Demolition of existing wall	-	H	2–3	2 days
	2. Lay the foundation for the new partition	1	H	2–3	3 days
	3. Build the partition wall up to the ceiling	2	H	2–3	3 days
	4. Install the wooden door	3	M	2	2 days
	5. Apply the first layer of stucco	4	M	2–3	2 days
	6. Apply the second layer of stucco	5	M	2–3	2 days
	7. Apply the first layer of paint	6	M	2–3	2 days
	8. Apply the second layer of paint	7	M	2–3	2 days
	9. Final inspection and clean-up	8	H	2–3	1 day
				TOTAL	19 days
3	1. Prepare the site	-	H	2–3	1 day
	2. Measure and mark the existing space	1	H	2	1 day
	2. Excavation for the new partition's foundation	2	H	2–3	2 days
	4. Building the new partition	3	H	4	4 days
	5. Install the wooden door	4	H	2	1 day
	6. Apply two layers of stucco to both sides of the partition	3	H	2	1 day
	7. Apply two layers of white latex paint to both sides of the partition	6	H	2	1 day
	8. Clean up the site	7	H	2	1 day
				TOTAL	12 days
4	1. Demolition of existing wall	-	H	2–3	1 day
	2. Excavation and preparation of foundation for new partition	1	H	2–3	1 day
	3. Installation of steel reinforcement for new partition	2	H	2–3	1 day
	4. Pouring of concrete for new partition	3	H	2–3	1 day
	5. Erection of new partition using concrete masonry units	4	H	2–3	2 days
	6. Installation of wooden door	5	H	1–2	1 day
	7. Plastering of new partition with first layer of stucco	5	H	1–2	1 day
	8. Plastering of new partition with second layer of stucco	7	H	1–2	1 day
	9. Painting of new partition with first layer of white latex paint	8	H	1–2	1 day
	10. Painting of new partition with second layer of white latex paint	9	H	1–2	1 day
	11. Clean-up and final inspection	10	H	1–2	1 day
				TOTAL	13 days
5	1. Preparation and planning	-	H	2–3	2 days
	2. Demolition of existing walls	1	H	2–3	2 days
	3. Construction of new partition	2	H	2–3	2 days
	4. Installation of wooden door	3	H	1–2	2 days
	5. Plastering	3	H	2–3	2 days
	6. Painting	5	H	2–3	2 days
	7. Clean-up and final inspection	6	H	2–3	2 days
				TOTAL	14 days

Table A1. *Cont.*

Participant	Task Name	Dependencies	Priority *	People Needed	Expected Duration
	1. Preparation and planning	-	H	2–3	2 days
	2. Demolition	1	H	2–3	3 days
	3. Masonry work	2	H	3–4	3 days
6	4. Door installation	3	H	2–3	2 days
	5. Plastering and painting	4	M	2–3	3 days
	6. Clean-up and final inspection	5	L	2–3	2 days
				TOTAL	15 days

* H = High, M = Medium, L = Low.

Appendix B

Figure A1 shows extracts from the interface with ChatGPT for some of the prompts used in this study.


(**a**)

(**b**)

Figure A1. *Cont.*

9. Applying Second Layer of Paint

- Task Name: Painting - Layer 2
- Task Priority: Low
- Task Dependencies: Painting - Layer 1
- Number of People Needed: 2
- Expected Duration of Task: 2 days

10. Final Inspection and Handover

- Task Name: Inspection and Handover
- Task Priority: High
- Task Dependencies: Painting - Layer 2
- Number of People Needed: 1
- Expected Duration of Task: 2 days

Please note that the expected durations of the tasks are based on the information provided, and actual durations may vary based on various factors such as the experience of the workers, the availability of tools and equipment, and any unforeseen issues that may arise during the project.

S   Add this information into columns

    Sure! Here is the information for the various tasks involved in this project organized in columns:

Task Name	Task Priority	Task Dependencies	Number of People Needed	Expected Duration of Task
Site Preparation	High	None	2	2 days
Procurement	High	Site Preparation	1	1 day
Masonry Work	High	Procurement	2	3 days
Door Installation	Medium	Masonry Work	2	2 days
Stucco Work - Layer 1	Medium	Door Installation	2	2 days
Stucco Work - Layer 2	Medium	Stucco Work - Layer 1	2	2 days
Surface Preparation	Low	Stucco Work - Layer 2	2	2 days
Painting - Layer 1	Low	Surface Preparation	2	2 days
Painting - Layer 2	Low	Painting - Layer 1	2	2 days
Inspection and Handover	High	Painting - Layer 2	1	2 days

(**c**)

Figure A1. Extracts of the ChatGPT interface with the prompts regarding (**a**) the initial instruction and the scope of work, (**b**) the request to detail tasks in a given format, and (**c**) displaying the proposed schedule in a table.

References

1. Chowdhary, K.R. Natural Language Processing. In *Fundamentals of Artificial Intelligence*; Chowdhary, K.R., Ed.; Springer: New Delhi, India, 2020; pp. 603–649. [CrossRef]
2. Xue, X.; Hou, Y.; Zhang, J. Automated Construction Contract Summarization Using Natural Language Processing and Deep Learning. In Proceedings of the 39th International Symposium on Automation and Robotics in Construction (ISARC 2022), Bogotá, Colombia, 13–15 July 2022; IAARC Publications. pp. 459–466. [CrossRef]
3. Di Giuda, G.M.; Locatelli, M.; Schievano, M.; Pellegrini, L.; Pattini, G.; Giana, P.E.; Seghezzi, E. Natural Language Processing for Information and Project Management. In *Digital Transformation of the Design, Construction and Management Processes of the Built Environment*; Daniotti, B., Gianinetto, M., Della Torre, S., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 95–102. [CrossRef]
4. Zou, Y.; Kiviniemi, A.; Jones, S.W. Retrieving similar cases for construction project risk management using Natural Language Processing techniques. *Autom. Constr.* **2017**, *80*, 66–76. [CrossRef]
5. Zhang, J.; El-Gohary, N.M. Integrating semantic NLP and logic reasoning into a unified system for fully-automated code checking. *Autom. Constr.* **2017**, *73*, 45–57. [CrossRef]
6. Tixier, A.J.-P.; Hallowell, M.R.; Rajagopalan, B.; Bowman, D. Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Autom. Constr.* **2016**, *62*, 45–56. [CrossRef]
7. Ding, Y.; Ma, J.; Luo, X. Applications of natural language processing in construction. *Autom. Constr.* **2022**, *136*, 104169. [CrossRef]

8. Hassan, F.U.; Le, T.; Lv, X. Addressing Legal and Contractual Matters in Construction Using Natural Language Processing: A Critical Review. *J. Constr. Eng. Manag.* **2021**, *147*, 03121004. [CrossRef]

9. Cho, J.; Lee, G. A Chatbot System for Construction Daily Report Information Management. In Proceedings of the 36th International Symposium on Automation and Robotics in Construction (ISARC 2019), Banff, AB, Canada, 21–24 May 2019; IAARC Publications: Waterloo, ON, Canada, 2019; pp. 429–437. [CrossRef]

10. Amer, F.; Jung, Y.; Golparvar-Fard, M. Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. *Autom. Constr.* **2021**, *132*, 103929. [CrossRef]

11. Naseem, U.; Razzak, I.; Khan, S.K.; Prasad, M. A Comprehensive Survey on Word Representation Models: From Classical to State-of-the-Art Word Representation Language Models. *Trans. Asian Low-Resour. Lang. Inf. Process.* **2021**, *20*, 74. [CrossRef]

12. Schomacker, T.; Tropmann-Frick, M. Language Representation Models: An Overview. *Entropy* **2021**, *23*, 1422. [CrossRef] [PubMed]

13. Locatelli, M.; Seghezzi, E.; Pellegrini, L.; Tagliabue, L.C.; Di Giuda, G.M. Exploring Natural Language Processing in Construction and Integration with Building Information Modeling: A Scientometric Analysis. *Buildings* **2021**, *11*, 583. [CrossRef]

14. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. *arXiv* **2020**. [CrossRef]

15. Ray, T. ChatGPT Is "Not Particularly Innovative," and "Nothing Revolutionary", Says Meta's Chief AI Scientist, ZDNET. 2023. Available online: https://www.zdnet.com/article/chatgpt-is-not-particularly-innovative-and-nothing-revolutionary-says-metas-chief-ai-scientist/ (accessed on 27 January 2023).

16. Korngiebel, D.M.; Mooney, S.D. Considering the possibilities and pitfalls of Generative Pre-trained Transformer 3 (GPT-3) in healthcare delivery. *NPJ Digit. Med.* **2021**, *4*, 93. [CrossRef] [PubMed]

17. Floridi, L.; Chiriatti, M. GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds Mach.* **2020**, *30*, 681–694. [CrossRef]

18. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv* **2019**. [CrossRef]

19. Hassan, H.A.M.; Marengo, E.; Nutt, W. A BERT-Based Model for Question Answering on Construction Incident Reports. In *Natural Language Processing and Information Systems*; Rosso, P., Basile, V., Martínez, R., Métais, E., Meziane, F., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 215–223. [CrossRef]

20. Moon, S.; Chi, S.; Im, S.-B. Automated detection of contractual risk clauses from construction specifications using bidirectional encoder representations from transformers (BERT). *Autom. Constr.* **2022**, *142*, 104465. [CrossRef]

21. Yao, D.; García de Soto, B. A corpus database for cybersecurity topic modeling in the construction industry. In Proceedings of the 40th International Symposium on Automation and Robotics in Construction (ISARC 2023), Chennai, India, 4–7 July 2023.

22. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *arXiv* **2020**. [CrossRef]

23. Jung, Y.; Fouad, A.; Golparvar-Fard, M. A systematic review on the requirements on BIM maturity and formal representation of sequencing knowledge for automated construction scheduling. In Proceedings of the 38th International Conference of CIB, Luxembourg, 11–15 October 2021; pp. 429–436. Available online: https://itc.scix.net/paper/w78-2021-paper-043 (accessed on 13 January 2023).

24. Aljebory, K.M.; QaisIssam, M. Developing AI Based Scheme for Project Planning by Expert Merging Revit and Primavera Software. In Proceedings of the 2019 16th International Multi-Conference on Systems, Signals & Devices (SSD), Istanbul, Turkey, 21–24 March 2019; pp. 404–412. [CrossRef]

25. Wang, Z.; Azar, E.R. BIM-based draft schedule generation in reinforced concrete-framed buildings. *Constr. Innov.* **2019**, *19*, 280–294. [CrossRef]

26. Hong, Y.; Xie, H.; Agapaki, E.; Brilakis, I. Graph-Based Automated Construction Scheduling without the Use of BIM. *J. Constr. Eng. Manag.* **2023**, *149*, 05022020. [CrossRef]

27. Awada, M.; Srour, F.J.; Srour, I.M. Data-Driven Machine Learning Approach to Integrate Field Submittals in Project Scheduling. *J. Manag. Eng.* **2021**, *37*, 04020104. [CrossRef]

28. Lin, J.J.; Golparvar-Fard, M. Visual Data and Predictive Analytics for Proactive Project Controls on Construction Sites. In *Advanced Computing Strategies for Engineering*; Smith, I.F.C., Domer, B., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 412–430. [CrossRef]

29. Hong, Y.; Xie, H.; Bhumbra, G.; Brilakis, I. Comparing Natural Language Processing Methods to Cluster Construction Schedules. *J. Constr. Eng. Manag.* **2021**, *147*, 04021136. [CrossRef]