<<DSAA-4693>>

Using AI to generate and iterate schedules based on scope documents - a world first

Gerard Cardoso, Damian Borowiec, Rhys Phillips, Leonie Anna Mueck, Dev Amratia

Abstract

In major construction projects the planning process is often labor-intensive and complex. From conceptualisation to project closeout, schedules must be managed and adjusted, often leading to bottlenecks that impede project flow. Project planners grapple with time-consuming manual inputs, frequent reworks, and the need to account for numerous interdependent activities in large and complex schedules. These challenges can result in scheduling delays, reduced visibility of key project issues, increased time to course-correct in projects, and increased costs, making efficient planning a persistent struggle in the construction industry.

This paper introduces Schedule Studio, a generative AI technology that transforms how construction schedules are created and kept updated. Leveraging the largest database of construction schedules globally and using plain text inputs, this tool automates the creation and iteration of detailed schedules, making the process faster, more flexible, and more impactful to project decision making. Key advancements in the fields of document processing and generative AI have made this innovation possible.

This paper also introduces a real life demonstration of this technology. This practical example illustrates how Schedule Generation not only streamlines planning but also has the potential to drive cross-disciplinary improvements, making it a transformative tool for modern construction planning. The trends in technology development and the reducing cost of such technologies will inevitably lead to its applications in the planning domain.

Given the current status quo in planning and the increasing maturity of schedule generation technology, the likely future of planning is one where the burden of manual reporting work is alleviated, schedules are routinely built using AI, and planners are freed to focus more on proactive horizon scanning, development of different delivery options, and the assessment of systemic risks.

Table of Contents

Abstract 1
Table of Contents 2
Introduction 3
Key Pain Points in Project Planning 4
Pre-FID 4
Construction 5
Project Closeout 6
Applying generative AI to scheduling and planning 7
Gathering and Synthesizing Project Requirements 7
From Documents to Insights: Challenges in Data Preparation 8
Generating Schedules 9
Increased Cognitive Capabilities - Parameters 9
Increased Cognitive Capabilities - Data 9
Expanded Context Windows 10
Navigating Structures through Language 10
Sequential Reasoning 10
Schedule Studio - From Scope Document to Schedule 11
How Schedule Studio solves points of friction in megaproject planning 15
Large Volumes of Data 15
Inherent Complexity and Decreasing Certainty 16
Time Constraints & Strategic Planning 16
Conclusion 17
References 18

Introduction

Construction projects, particularly the larger 'mega' and now 'giga' projects, are predicted to represent $9 trillion of spending in 2025 [1]. Delivering them successfully requires teams and plans of enormous complexity. At the same time, they are of rising global importance as they shape and reshape the fabric of society. For example, the successful deliveries of mega and giga-projects in infrastructure are crucial to reducing carbon emissions. Spending on infrastructure represents roughly 14% of global GDP [2] and yet the record of project delivery has been consistently below expectations with cost and schedule overruns nearly ubiquitous [3] and attempts to curb this routinely falling short [4].

The root causes of the generally accepted record of project overruns have been explored in detail. The range of biases that influence human judgement both consciously and subconsciously [5] and the nature of projects as complex systems [6] are both well known and will have been experienced by anyone familiar with major infrastructure projects. Planning and scheduling are inextricably linked to this problem as the project schedule that the planning function develops, and ultimately owns (though effective planning requires input from all teams), is the key contractual document that informs delivery strategy as well as progress reporting, delay claims, and payment when key milestones are reached.

However, producing high quality schedules and then maintaining them is a substantial burden, especially in light of increasing demand for critical infrastructure and an aging workforce [7]. In particular, the following pain points are prevalent in the planning community:

1. Very large quantities of diverse data need to be sourced, understood, interpreted and actioned on in very short timeframes.
2. Certainty must naturally decrease further into the future but projects are necessarily planned as though the future is knowable. The difficulty with larger projects is exacerbated greatly by the complexity inherent in tens of thousands of lines of schedule.
3. The time constraints and complexity mean that it is exceedingly difficult to look forward strategically and plan for different delivery scenarios and to effectively learn lessons to inform future projects.

With the great improvements made in the field of artificial intelligence (AI) and the emergence of mature AI solutions for related problems such as risk analysis [8, 9, 10], the field of planning is primed to take advantage of this technology, supported by their increasing acceptance as a commonplace tool in infrastructure projects, and the pressures on the industry to deliver more, faster, while not sacrificing quality or cost. Generative AI is a particularly exciting opportunity here as it allows not only for predicting the outcomes of plans based on past data but also enables the generation of documents relevant to planning, risk management and project management. Since the launch of ChatGPT, a chatbot based on Large Language Models (LLMs) [11], these capabilities have been widely known in the community and have been integrated in various products used in the industry.

Furthermore, information in the construction industry is increasingly becoming digitized and shared in real time across platforms like Procore, SharePoint, and Microsoft Teams. This interconnected ecosystem has greatly enhanced collaboration among project stakeholders, offering streamlined document-sharing workflows and continuous visibility into project updates. Virtual meetings and email communications are now routinely captured and stored, ensuring that critical information is accessible for future reference and integration with other applications. Such practices not only enhance transparency but also provide a comprehensive repository of project data that can be leveraged for advanced tasks like planning and scheduling.

These practices represent an opportunity for generative AI methods to enhance productivity within the construction industry. Recognizing the value of integrating with existing digital ecosystems, LLM providers are actively developing solutions that seamlessly connect with various data sources. This focus on integration allows LLMs to access pertinent context, enabling them to generate more accurate and personalized outputs. As these models continue to evolve, their ability to embed within users' workspaces and provide tailored, specific answers will become a key differentiator in the market. However, even with these advances in collaboration technology, creating and maintaining comprehensive construction schedules - from conceptualization to project closeout - remains a challenging and time-consuming endeavor.

In particular, producing schedules for construction projects has so far been impossible with generative AI. Schedules are large and complicated data structures represented by graphs with activities and logical connections between these activities and milestones. Other information contained in a schedule are work breakdown structures, task lengths, lags etc. making a construction schedule a complicated amalgam of structured and unstructured data. This is what has so far made the generation of schedules impossible with generative AI. In this paper, a solution to this problem is presented: Schedule Studio is a generative AI system with the ability to produce and iterate on complex project schedules. The paper will first describe in detail how the status quo in planning and scheduling of major infrastructure projects struggles to meet the requirements of the modern world. It will then discuss the fundamental underpinnings of cutting edge AI technologies to understand the convergence of factors that make LLM-based schedule generation possible. After presenting the solution developed by the authors, the implications for how projects will be delivered in the future is explored.

Key Pain Points in Project Planning

The planning and scheduling functions have been under particular pressure in recent years within mega and giga-projects as well as in smaller projects. Throughout the life cycle of a project the project planner (or planning function) is the custodian of the schedule, responsible for developing and iterating on the schedule in line with contractual requirements. Most projects, whether large or small, require a monthly periodic update to the schedule that reflects the actualisation of tasks, the changes to schedule logic (including additional or dropped scope), and any resulting change in the logic driven dates of key milestones. This process is defined in the organisation's planning handbook and must comply with the project contract. On smaller projects there may be a single planner covering a number of projects in a programme, whereas on large projects there may be an entire team of planners for each section of the project.

Regardless of scope, scale, or size, the problems faced by planners every day are similar. The following key pain points in the planning process have emerged as bottlenecks to improving project deliverability and elevating the planning profession within the project delivery community. This section will step through the high level lifecycle of major infrastructure projects and highlight key issues at each stage.

It should be stated at the outset that it is not the position of the authors that AI will replace the planning function in megaproject delivery. Rather, the introduction of AI Schedule Generation will serve to elevate the planning function and unlock greater strategic value from experienced project professionals.

Pre-FID

Prior to the Final Investment Decision (FID) the primary issues faced by planners are threefold:

1. The tight timelines mandated and the sheer volume of information to be obtained, understood, and then integrated into the resulting plan.

The pre-FID stage requires the synthesis of diverse inputs, including engineering specifications, procurement strategies, environmental assessments, regulatory compliance, stakeholder requirements, and financial models. The interdependence and dynamic nature of these inputs present significant complexity, often compounded by strict deadlines. The compressed timelines typical of the pre-FID phase place immense pressure on planners to deliver comprehensive and actionable schedules. The speed at which data must be obtained, analyzed, and incorporated into the project plan often limits the opportunity for thorough validation and iterative refinement. As a result, there is an elevated risk of incomplete or misaligned data impacting the accuracy of early project plans.

It is not unusual for tender teams to have to access, interrogate, understand, and then integrate thousands of pages of complex documents [12] and then formulate a coherent delivery strategy that both meets tender requirements in terms of time and cost, but also does not expose the contractor to undue levels of risk.

2. The reality that certainty must decrease further into the future, yet multi-year projects require detailed plans with hard dates for key milestones up front.

The challenge of declining certainty over time is a central concern for even expert planners. The further a project looks into the future, the greater the difficulty in predicting key variables, including material costs, resource availability, labor markets, regulatory requirements, and stakeholder expectations, without considering wider macroeconomic events as have been experienced in the last 5 years. This inherent uncertainty undermines the feasibility of generating a detailed and reliable schedule that extends years into the future.

At this stage, planners must acknowledge the limitations of deterministic scheduling approaches. Attempting to lock in a detailed schedule based on assumptions that are likely to change can create unrealistic expectations among stakeholders and lock the project into an inflexible framework. Such rigidity increases the likelihood of costly delays and rework when unforeseen changes inevitably occur later on.

One approach to overcoming this inherent uncertainty problem is the rolling wave planning approach. This involves creating detailed schedules only for near-term activities while leaving long-term tasks at a higher level of abstraction. This iterative approach allows planners to refine details as more information becomes available, ensuring schedules remain realistic and actionable. However, when contracts require key milestone dates to be agreed in the baseline schedule, projects are still at risk of baking in significant risk and uncertainty as to their deliverability by not building out a more detailed plan up front. Additionally, where projects are access driven - for example rail projects requiring possessions or transmission projects requiring energisation - this problem becomes fractal in nature, with multiple hard milestones to plan for and inadequate certainty as to the deliverability of the required preparatory works.

3. The shifting nature of demand - including renewable energy, nuclear, and data centres - means that contractors are shifting strategies to delivering work they are less familiar with.

An emerging challenge for many contractors is the increase in demand for projects that the organisation may be unfamiliar with delivering, including those in nuclear (including Small Modular Reactor (SMR) projects), renewable energy, and data centre spaces. Some contractors have already started to pivot their strategy and pin their future profitability on the growth of these markets [13].

This shift in demand acts as a force multiplier for the previously explored problems of tight tender timelines and the rigidity of multi-year projects with teams forced to reckon with more unfamiliar, and hence riskier, project proposals.

Construction

During construction, the schedule that has been agreed upon during the tendering and contract award process is established as a baseline and then delivered against. Here, the planning discipline can be broadly divided into two camps:

1. Operational planners: focused on the day to day process to collect data, update the schedule, and generate reports. This role is characterised by executing the same processes each month.

In a highly process driven role, the points of friction in execution arise when information must be retrieved, integrated, and analyzed within a monthly cycle that imposes strict time pressures. Similar to the risk management process [8] the time pressure of each period within a project’s execution means that a majority of an operational planner’s time can be spent feeding information into reporting systems, rather than actioning this information.

The problems are exacerbated on larger programmes where multiple schedules may be managed independently or need integrating to give project leadership an overarching picture of progress and risk. The process of managing and integrating project schedules is highly labour intensive, requires deep experience and expertise, and is prone to error.

Even if the labour and experience challenges can be overcome there is still the issue of the resulting integrity of the outputs. There are many tools - both proprietary and in-house - that analyse the integrity of a schedule and it is widely recognised that better schedule integrity is correlated with better project outcomes (though the degree of causality is debated) [14, 15]. Without strong schedule integrity the results of any schedule analysis - even a simple critical path analysis that provides project leadership with insights into key drivers of completion - is fatally undermined.

2. Strategic planners: focused on the deliverability of the project and how potential threats can be mitigated and avoided. This role is generally more free-form and less constrained by a defined process, however it is still fundamentally constrained by the nature of plans, and planning and cadence based project execution systems.

Strategic planning is generally undertaken on larger megaprojects where the potential for exogenous shocks and emergent properties from the complex nature of the project schedule make it necessary to engage in strategic foresight. Complexity is a concept that humans struggle with, often being confused for the more commonplace 'complicated', and it is this complexity of large schedules, meaning that the knock on effects of delays are almost impossible to foresee manually, that makes strategic planning so necessary.

As projects are progressing, the task of exploring schedule and delivery options and developing more detailed plans for later stages (either when using the rolling wave planning methodology or when developing commissioning or turnaround plans) falls to the strategic planning function. However, the sheer size of this task often means that good intentions fail to deliver. When combining rigid contractual requirements with rigid deterministic schedules that are routinely tens of thousands of lines long (one schedule the authors recently analysed was over 140,000 lines), the ability of human planning teams to explore schedule options is intrinsically limited. Rather than being a continuous process of exploring the solution space of delivery, delivery scenarios are more routinely explored at pre-defined phase gates which limits the utility of the exercise.

At critical junctures of the project, such as rebaselining exercises, the exigent circumstances demand a detailed response in the form of an exploration of schedule options that are truly deliverable. Projects in delay are under intense time and cost pressure and so at the time when the widest variety of delivery options should be explored, projects are still constrained by the usual time, cost, and contractual pressures.

The ability to quickly and efficiently react to the need to develop a new reality and generate a series of feasible delivery solutions while respecting existing constraints is a key challenge within strategic planning. Attempts have been made within the academic sphere to address this challenge, for example, with Dynamic Adaptive Planning (DAP) within the Decision Making under Deep Uncertainty (DMDU) framework [16], but thus far this approach has not become mainstream due to the infeasibility of developing a pareto front of schedule options in sufficient detail.

Project Closeout

Project teams are transient entities that are constituted and then dispersed approximately every two to four years. This makes learning lessons from a wide range of projects hard, and applying them to new projects even harder [17]. Lessons learned exercises attempt to capture the learnings from each project that the organisation feels are pertinent for future projects but are inherently limited by the flux of personnel across the life of the project, the diminishing quality of memory over time when looking back from the end of a project (when lessons learned exercises are generally carried out), and the biases that influence human perception which skew how one views events, and therefore how one learns from them.

While the volume of physical work to be executed during project closeout is lower than during the construction phase, even for projects with detailed commissioning sequences, the volume of administrative work may actually increase at this time with the burden of compiling ‘as built’ and handover documentation exercises being placed on a smaller team. The competing priorities at this stage of the project means that lessons learned exercises can be performative and lack efficacy for large organisations.

Applying generative AI to scheduling and planning

The pace of evolution in AI technology directly relevant to the problem of building and updating construction project schedules is immense. In the past two years, the field has advanced from Language Models that were unable to answer simple questions correctly with any degree of reliability (and regularly hallucinated answers rather than admitting a lack of knowledge) to a point where the most advanced generative AI is able to tackle complex, hierarchical, and multi-level reasoning tasks with relative ease.

It should also be noted that the cost of this intelligence is declining rapidly. In just the last two years the cost per token of GPT4 level intelligence has decreased by a factor of 240 [18]. This means that not only can generative AI be used to complete tasks that were previously thought to be exclusively the reserve of experienced knowledge workers, but also these models are able to do so for a rapidly decreasing cost, making this technology potentially affordable for projects and organisations of all sizes.

The rapid advance of AI technology and the equally rapid decline in its cost makes it inevitable that these technologies will come to be applied to the complex task of project planning. In this section, the main capabilities and functionalities that are necessary to build such a generative AI planning system are laid out.

Gathering and Synthesizing Project Requirements

Planning in construction project management often hinges on the ability to interpret extensive, multi-format documentation and translate it into actionable guidelines. As discussed in the previous section, planners must reconcile diverse inputs - ranging from design blueprints and regulatory policies to contracts and resource constraints - while ensuring that final deliverables meet both deadlines and quality standards. Traditionally, this has been a time-consuming endeavor, demanding meticulous reading, cross-referencing, and distillation of technical details.

Increasingly, Large Language Models (LLMs) offer new avenues for automating parts of this synthesis. While models like OpenAI’s GPT-4 [19] and Google’s Gemini [20] demonstrate remarkable skill in understanding and generating text, their out-of-the-box usefulness for construction projects can be limited by the absence of project-specific knowledge. As a result, while they can provide generalized insights, they struggle to address nuanced queries that require proprietary data or involve projects initiated after their training cutoff date.

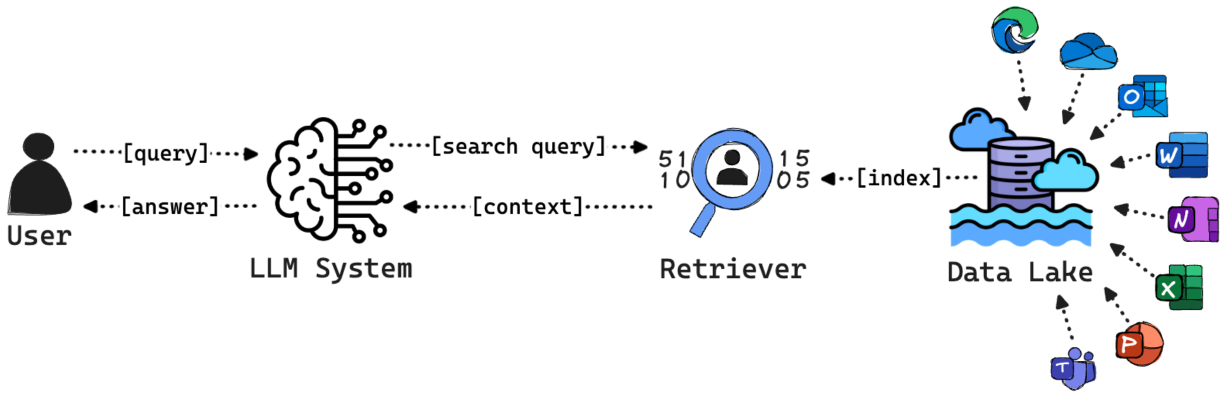


Figure 1. An illustration of Retrieval Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) has emerged as a key strategy to address this gap, enabling LLMs to incorporate proprietary or recently updated information [21]. RAG operates by augmenting model prompts with relevant passages from project documents at query time. For example, an LLM prompting workflow might:

1. Index documents (text, images, or structured data) from project archives (i.e. a data lake)
2. Identify the most pertinent sections (e.g., snippets detailing building codes or a line item from a resource plan) utilising retrieval methods [21]
3. Feed this data into the LLM as context alongside the instruction prompt

This workflow is illustrated in Figure 1. RAG solutions can substantially improve the accuracy and relevance of LLM-generated insights - an essential capability for building detailed schedules that accurately reflect current project conditions. Consequently, building an efficient retrieval system becomes crucial to identify and deliver the most pertinent data to the generative task at hand - e.g. generating the Work Breakdown Structure (WBS).

From Documents to Insights: Challenges in Data Preparation

Modern construction projects generate and rely on a vast array of documents - ranging from text-heavy specifications and contracts to intricate blueprints, tables, and annotated diagrams. For AI-driven schedule generation to be truly effective, these documents must be converted into a format that LLMs can process and understand. Achieving this involves overcoming several challenges, from accurately parsing complex layouts to understanding and translating diagrams and drawings.

One of the first steps in designing a RAG workflow is extracting text from documents with heterogeneous structures. Although Optical Character Recognition (OCR) is a common tool for digitizing paper or PDF documents, it often struggles with:

- Complex tabular data in resource estimates or work breakdown structures
- Visual elements that can't be captured purely as text - e.g. diagrams, blueprints
- Handwritten notes or layered comments that require more advanced recognition

Recent developments in AI have facilitated more sophisticated approaches beyond traditional OCR. Vision Language Models (VLMs), trained on both text and visual features, can interpret document structures (headings, columns, image boundaries) and preserve relationships between different elements [22]. Additionally, these multi-modal[1] models integrate visual and textual inputs - enabling them to understand and extract key information from images or diagrams directly, which is particularly advantageous in construction projects where visual data often carries critical information.

[1] multi-modal means the model accepts multiple modalities such as text, images, video etc.

The advancements in document understanding and retrieval (RAG) technologies, combined with the capabilities of LLMs, address a critical need in project scheduling: the ability to efficiently integrate information from diverse sources. Construction projects often involve a wide array of documents, all of which must be synthesized, analysed and integrated to create accurate and realistic schedules. Without the aforementioned technologies, the process of extracting, structuring, and utilizing this information would remain fragmented and inefficient, limiting the potential of LLMs to generate meaningful outputs. This integration of cutting-edge AI ensures that project schedules are informed by the right data, supporting better decision-making and more effective project management.

Generating Schedules

Construction project schedules are very complex data structures. A typical schedule is composed around key phases (planning & design, procurement, construction, handover, closeout etc.) and milestones (drywall installed, final permits etc.), guiding the project from inception to completion. Furthermore, schedules are commonly broken down into a Work Breakdown Structures (WBS) - a hierarchy of progressively more granular deliverables, work packages and finally, thousands of individual project activities. Alongside its name, each such activity has further information and meta-data associated with it. Finally, all of the schedule's activities are connected together, encoding their intricate interdependencies that dictate project activity ordering and scheduling.

In a general sense, LLMs operate purely on text, ingesting input tokens to then generate output tokens [23] (essentially taking a text input and producing a text output). It is through this process that higher-level artificial intelligence capability emerges (e.g. question answering, report generation, conversing with humans and more) [24]. As such, it is challenging to even imagine how such a purely text-based system could be used to generate construction project schedules - complex amalgamations of both structured and unstructured data. However, over the past several years, new developments in Large Language Models (LLMs) and the supporting technologies discussed in the previous section have created fertile ground for automating the generation of these project schedules from raw source documents, design files, communications, and other data. Broadly, five major areas of advancement have been pivotal:

<u>Increased Cognitive Capabilities - Parameters</u>

The core of knowledge (concept associations) and capabilities possessed by LLMs is encoded within their parameters [25, 26]. LLMs are referred to as "Large" Language Models precisely because of an increased number of these parameters compared to any other models in Machine Learning. Since 2017, parameter count in Language Models has grown from around 350 million (BERT [27]) to nearly 2 trillion (OpenAI GPT4) [28] - an exponential jump in scale [29]. A larger parameter space allows the model to encode more nuanced information acquired during training, including more nuanced associations between concepts. This in turn, when combined with a large enough set of training data, allows higher-level cognitive capabilities to emerge - capabilities essential for a task as complex as project planning.

<u>Increased Cognitive Capabilities - Data</u>

To "fill up" these parameters with concepts, and to teach the models to convey them through language, LLMs are trained by being repeatedly presented with vast amounts of diverse text obtained from various sources [30]. Since 2017, the size of training datasets has increased from 3.3 billion tokens (BERT) or 2.5 billion words, to 13 trillion tokens (OpenAI GPT4) [28, 31]. Models have increased in size as well to the point where the largest LLMs contain compressed within them, nearly the entire publicly available knowledge from the internet, as well as books, textbooks, proprietary data sources, online conversations from forums and more [32]. In large, complex construction projects, it is common for stakeholders to come from diverse backgrounds - engineers, architects, legal teams, and environmental specialists, among others [33]. Modern LLMs are often equipped with baseline knowledge in all of these areas and can help fill in gaps where a single discipline's expertise may not suffice [34].

In the realm of schedule generation, these exponential improvements translate into improved ability to, for example, suggest logical dependencies between schedule activities and spot potential conflicts or logical errors early during schedule conceptualisation. At the same time, by accommodating more nuanced contexts, larger models can lend support across multiple project phases. Whether it is the design phase, construction phase, or handover, the model is equipped to handle distinct sets of requirements without losing fidelity or relevance.

Expanded Context Windows

A common practice when working with LLMs is to include additional information in the "prompt" during generation. The model can leverage this information to then answer queries more factually and focus on the topic better, expanding its knowledge beyond that embedded in its parameters [35]. The amount of information that can be provided during query time by the user is known as the context window. In early 2023, LLM context windows maxed out at 4096 tokens (around 8 to 10 pages of A4 text) [36]. In as little as 2 years, this context window increased to over 1M tokens (with experimental models surpassing 10M) [37]. This larger context capacity means that entire design documents, building codes, contracts, textbooks and other project documentation can now be provided to the model for consideration during schedule generation. By ingesting such comprehensive data, an LLM can form a more holistic understanding of the project's requirements, constraints, and objectives. Additionally, by grounding the generation context in the aforementioned sources of truth, LLMs are much less likely to "hallucinate" or fabricate information.

Navigating Structures through Language

The improvements in parameter size, diversity of training data and increased context windows in state-of-the-art LLMs also resulted in enhancement of ability to interpret, navigate and reason about structured data [38]. Traditionally, language models were seen as advanced text parsers; however, the newer generation of LLMs has shown remarkable progress in modelling hierarchical and interdependent structures [39]. Construction schedules can be conceptualised as large, interconnected network graphs of tasks, milestones, and dependencies. Whereas older models might only handle linear text, the newest LLMs can encode and interpret various types of structured information (tables, markup languages such as Markdown, JSON, YAML, XML and others) [40, 41]. As such, schedule graphs could be represented in various ways to become inputs and outputs of modern LLMs.

However, graphs are not the only information structures embedded in project schedules. A WBS naturally forms a tree-like hierarchy of deliverables, packages, and activities. While early LLMs struggled with reasoning over such larger structures, recent models are better at hierarchical reasoning, which could allow them to comprehend how high-level phases decompose into more detailed levels and how to differentiate between them when generating parts of schedules [42].

By navigating complex schedule structures in an iterative, context-aware manner, next-generation LLMs can help to bridge the gap between raw project inputs (such as design documents or contract clauses) and the final, detailed project plan.

Sequential Reasoning

Construction project planning and scheduling are inherently complex processes that require advanced reasoning due to the dynamic and interdependent nature of tasks involved. Unlike other types of project management, construction planning must consider a multitude of variables, including task dependencies, resource allocation, regulatory compliance, safety requirements, and time constraints.

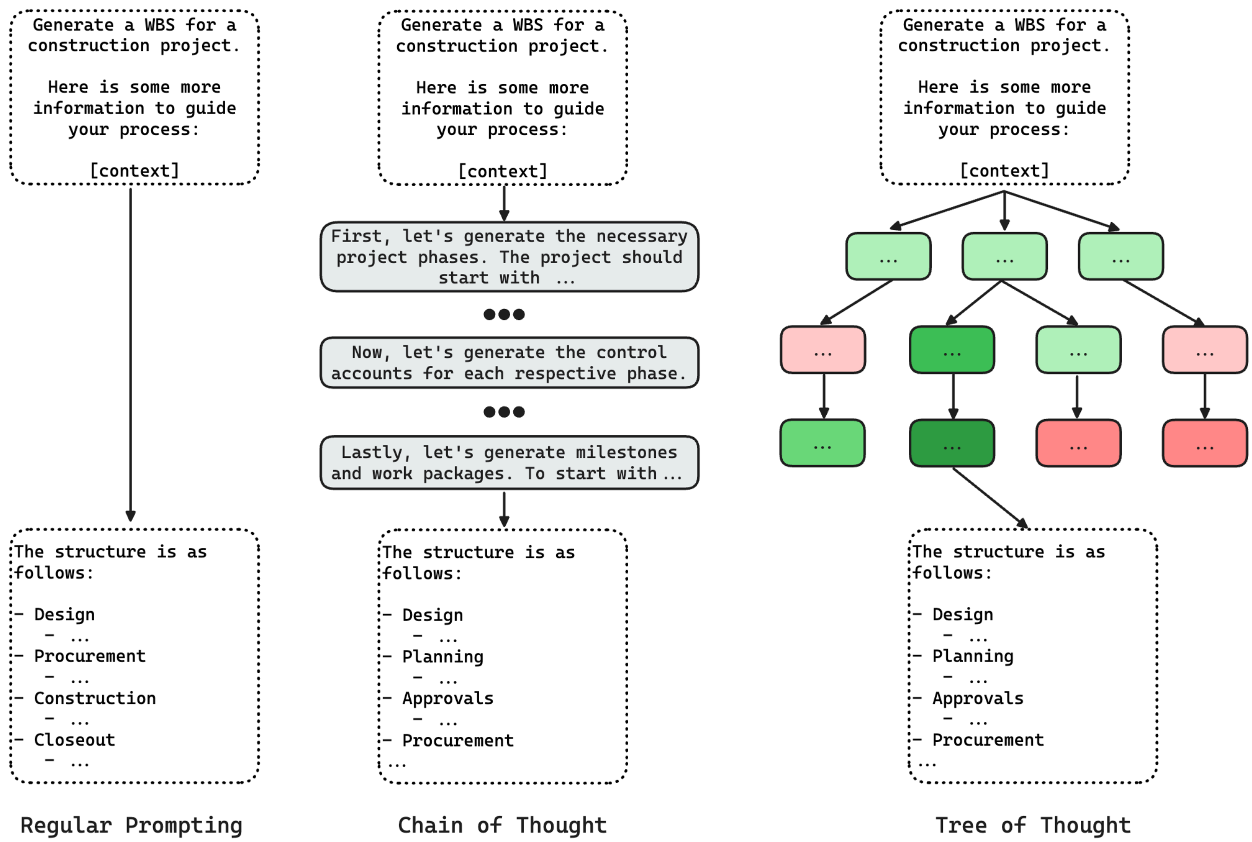


Figure 2. Prompting LLMs to think about the query before responding improves their accuracy and reasoning capabilities

Early LLMs were constrained to generating immediate answers to queries posed by users by their parameters, training data and context windows, leaving little room for "thinking through" the problem. As improvements in these aspects continued to be made, state-of-the-art LLMs became much better reasoning machines. To attempt to address this, techniques such as Chain-of-Thought (CoT) prompting have been developed to encourage LLMs to think in a step-by-step manner before arriving at an answer, simply by asking them to: "think step by step" [42]. This greatly improved their reasoning ability [42, 43]. Building upon these methods, other, more complex forms of prompting approaches such as Tree-of-Thought [42] have been developed where the LLM can explore multiple potential reasoning paths formed from these aforementioned thinking steps, backtracking up the tree of thought if a given reasoning path did not arrive at a suitable solution to the problem.

All of these advancements have recently culminated at novel reasoning LLMs such as O1 [44] from OpenAI or DeepSeek R1 [45]. These models have been proactively trained using Reinforcement Learning to break down complex tasks into smaller sub-problems which are then tackled independently. Once this "thinking" process completes, the final answer is generated and returned back to the user. In terms of applicability to automated schedule generation, for example, when creating a schedule for a multi-phase project like a commercial building, these models can determine a logical sequence for tasks such as site preparation, foundation laying, structural assembly, and interior finishing, while accounting for critical path considerations, overall project duration or other constraints.

Schedule Studio - From Scope Document to Schedule

The technological advancements described in the previous section have made it possible to build a solution, which in this paper is referred to as Schedule Studio, that ingests a request for proposal or scope document and outputs a realistic schedule. Schedule Studio combines the technology described above and makes use of previous work in the schedule risk analysis domain [9].

It should be noted that while the technology underpinning AI schedule generation is mature, its application to construction project schedule generation is not and so this solution - the first of its kind published anywhere in the world to the knowledge of the authors - is necessarily an internal experiment rather than an example of a real project being delivered with an AI generated schedule. The delivery of a real-life project with a schedule initially developed or continually updated by AI is expected to be at least two years away given current technical maturity.

Schedule Studio works by taking in documents commonly used at the start of projects to develop schedules - for example RFPs, scope documents - and generates a logically linked schedule in roughly 10 minutes.

One of the key developments made in this research is to develop a construction project specific fact extraction process that can ingest documents several thousand pages long and identify and classify relevant facts and datapoints that will be pertinent for developing the schedule. This allows the system to rapidly distill the information germane to the process of building the schedule and collate it in one place.

In Figure 3 (below) an example of this fact extraction process is demonstrated using a real RFP from an Interstate upgrade project from 2023. This document is several hundred pages long and contains text, tables, images, and tickboxes. While previous generations of Language Models were unable to handle such varied data structures and such document lengths it has been found that the Schedule Studio platform is able to accurately extract all pertinent information that human reviewers were also able to extract. However, this fact extraction process takes place within the 10 minutes it took to build the resulting schedule end to end, far faster than any human could digest such a large volume of information.

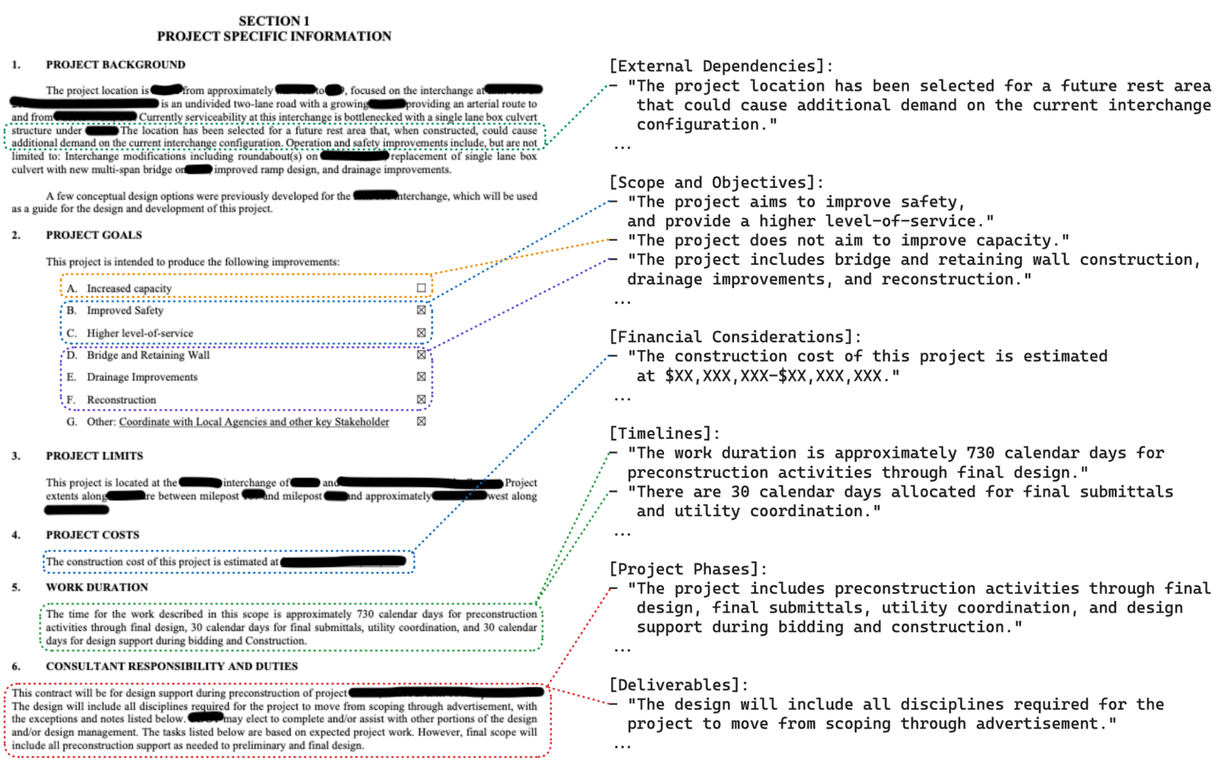


Figure 3. Using Vision Language Models to extract factual information from a reference document

Filtering out only the salient information from documents ensures that any downstream LLM calls will not suffer from exceeding the context length and also that the models are focusing on the right information to build out the schedule.

Once the fact extraction process is complete, the authors also found benefit in using an LLM to generate a summary of the overall project. This is useful in seeding the LLM to understand what the overall project is about in a few sentences. From the 100 page document above, a summary was created which is shown in Figure 4. The short summary is subsequently fed into all other LLM calls to improve the quality of, for example, the WBS generation and the project sequencing.

```
The I-25A at Exit 108 Improvements project is a transportation infrastructure project located in
Pueblo County, Colorado. The primary objectives of the project are to improve safety, and provide
a higher level-of-service at the I-25A and Purcell Blvd interchange.

The key components of the project scope include:

    -   Interchange modifications, such as replacing a single lane box culvert with a new
        multi-span bridge and improving ramp design
    -   Drainage improvements, including analyzing the existing hydrology and sizing new drainage
        facilities
    -   Roadway improvements along Purcell Blvd and connecting appurtenances to the I-25A
        improvements
    -   The project will be delivered through a consultant contract with CDOT overseeing the
        environmental clearances, design, and construction. The consultant will be responsible for
        tasks such as data collection, field investigations, utility coordination, drainage
        design, traffic analysis, and preparing construction plans and specifications.

The project is estimated to cost between $28-35 million and is expected to take approximately
1.5-3 years to complete the preconstruction activities through final design. Key milestones
include an initial project kick-off meeting, periodic progress meetings, public meetings, and a
Final Office Review (FOR) prior to finalizing the design.

The project will require coordination with various stakeholders, including the City of Pueblo,
Pueblo West Metropolitan District, Pueblo County, Pueblo Area Council of Government (PACOG),
Federal Highway Administration (FHWA), and Colorado Department of Public Health and Environment
(CDPHE). Environmental analysis and documentation under the National Environmental Policy Act
(NEPA) will also be a critical component.

Overall, the I-25A at Exit 108 Improvements project aims to enhance transportation infrastructure
and operations in the Pueblo County region through a comprehensive set of interchange, roadway,
and drainage improvements.
```

Figure 4. A generated summary made by Schedule Studio from the reference document provided

Once the project context has been established, the next step in the Schedule Studio pipeline is to generate a comprehensive WBS that captures the breadth of work and scope in the project. Figure 5 shows a subset of the WBS generated for this example Interstate project. Schedule Studio will iteratively develop down the WBS hierarchy with the goal of capturing all the necessary work that needs to be carried out from pre-construction all the way to project close-out. LLMs are quite effective at this task because of the amount of domain specific knowledge that they have been trained on from the internet.

This is a natural step to invite comments and feedback from the planner-in-the-loop, before Schedule Studio continues generating the schedule based on the information it has created thus far. The summary and facts provide an interesting opportunity as an interface between the planner and the AI - the authors believe that the project summary, project facts, schedule assumptions are a suitable level of abstraction for planners to provide feedback and for the AI model to incorporate this feedback for future generations. Editing the source documents would be more prone to being ignored in the synthesizing process, while providing feedback at the granular schedule level is too low-level and time consuming for the planner.

Across several realistic trials of Schedule Studio, the authors found that LLMs are more prone to add more detail in the generated WBSs at the pre-construction and project close-out phases. This could indicate a bias in the LLMs where they have been exposed to much more of these kinds of activities than more technical, construction & engineering content. This opens up a future research avenue to fine-tune models on more technical engineering domain content to improve their reasoning capabilities in this area of project planning. This has shown great success in other domains such as mathematical reasoning, and coding [46].

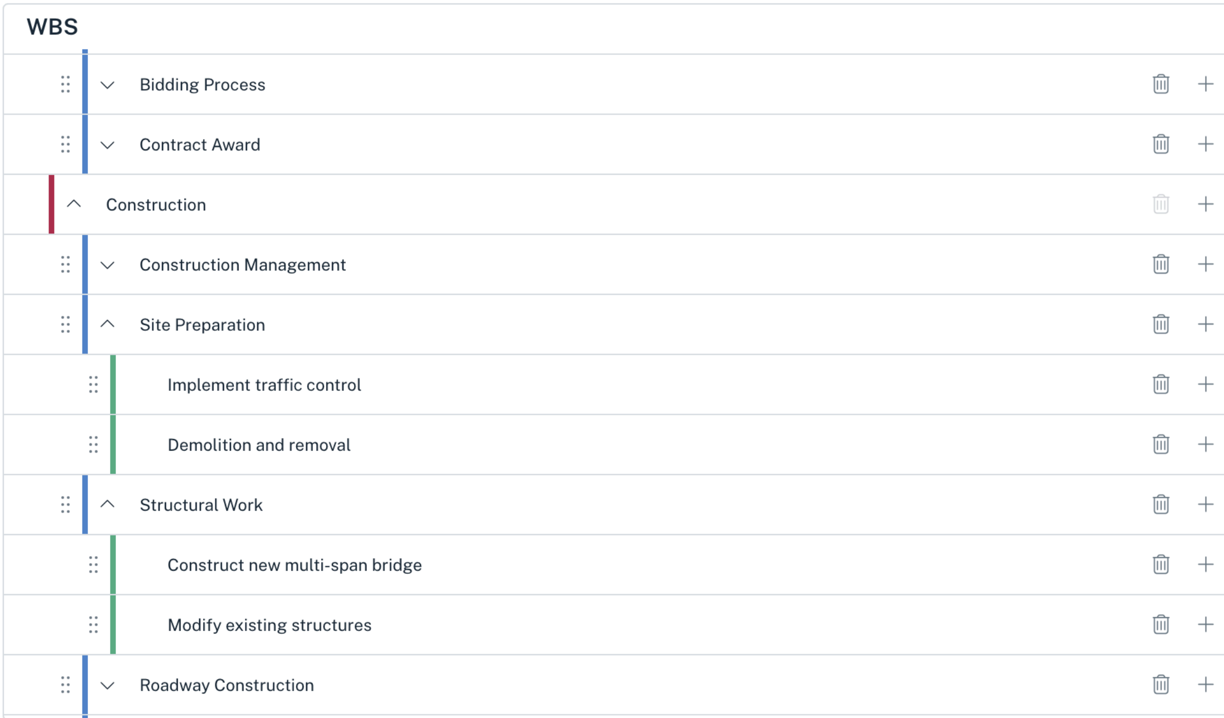


Figure 5[2]. A Work Breakdown Structure (WBS) generated by Schedule Studio for the example Interstate project

The lowest level of the WBS tree generated can then be fed into the next step of the Schedule Studio pipeline, which is all about how to sequence the work on the project. This step is significantly more complex for an LLM as it requires reasoning about what areas of the project have direct / indirect dependencies, to understand what needs to be done in sequence or what can be done in parallel. The current version of Schedule Studio also does not consider resourcing constraints which is an important consideration when determining how much work can be parallelised. This is discussed further in the future work section.

Furthermore, LLMs need to be able to generate exact graph logic by outputting all the relationships between all activities in the conceptual schedule. This can be difficult for these models to do since they were trained and fine-tuned to output natural language rather than graph-like logic. However, larger models like GPT-4o and Claude 3.5 Sonnet show remarkable ability to adapt to new kinds of syntax necessary for sequencing a project.

Figure 6 highlights a section of the conceptual schedule generated for the Interstate project. The activities in the schedule are the lowest level components of the WBS, and the LLM has determined in what order these activities should be done. This particular step of the pipeline is perhaps the most important to get right as it has implications on forecasting dates of key milestones on the project. Also crucial in this area is how to populate the schedule with realistic durations for each activity, which requires a lot of reasoning about the type of work being done and in what context (e.g. site conditions / amount of material being installed). More work is needed in this area to make

[2] Figures 5 and 6 were created using nPlan's Schedule Studio software

this kind of product reliable across industries. Suggestions for how to do this are discussed in the future work section.

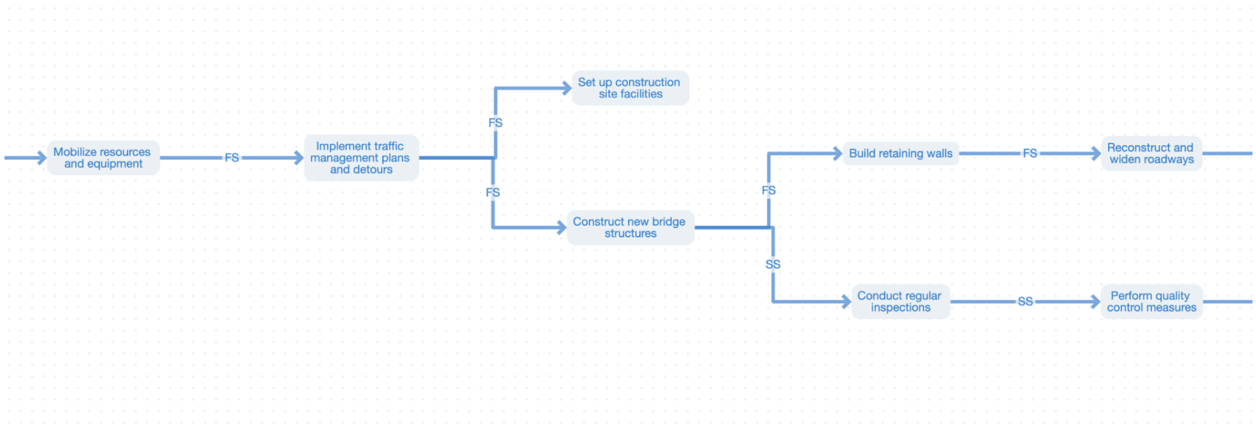


Figure 6. A subsection of a generated schedule from Schedule Studio for the example Interstate project

In summary, the Schedule Studio pipeline has three core components to take project documentation and generate a conceptual schedule. The first step, leveraging OCR and VLMs, extracts the most salient information about the project, as well as builds useful summaries of the project to increase the quality of LLM outputs in the rest of the pipeline. Secondly, a WBS for the project is created iteratively which provides the first insight into how the LLMs are reasoning about what needs to be delivered on the project and how that may be done. This is a natural opportunity for planners to get involved in critiquing outputs to align subsequent generations to their expectations. Lastly, these key activities can be sequenced to output a first draft of the project schedule in roughly 10 minutes. Given the speed of this generation, one could explore alternative scenarios of the project delivery, prompting the model to sequence the project differently and end up with a more considered final version of the project plan.

How Schedule Studio solves points of friction in megaproject planning

The major points of friction and inefficiency in the planning function of major projects have been summarised above as:

1. Very large quantities of diverse data need to be sourced, understood, interpreted and actioned on in very short timeframes.
2. Certainty must naturally decrease further into the future but projects are necessarily planned as though the future is knowable. The difficulty with larger projects is exacerbated greatly by the complexity inherent in tens of thousands of lines of schedule.
3. The time constraints and inherent complexity mean that it is exceedingly difficult to look forward strategically and plan for different delivery scenarios and to effectively learn lessons to inform future projects.

Tools such as Schedule Studio can help address these pain points which could, in turn, lead to a quantum leap in improving productivity and efficiency on mega and giga-construction projects. In the following sections, the ramifications of generative AI solutions to the major points of friction in planning are discussed in more detail.

Large Volumes of Data

The problem of the sheer volume of information which must be accessed and understood in order to apply it to the task of generating a construction schedule is perhaps the simplest task for generative AI to provide a solution to. Schedule Studio demonstrates that the advances in Language Models, Retrieval Augmented Generation (RAG), Optical Character Recognition (OCR) and Chain-of-Thought-like model reasoning have the potential to improve this pain point dramatically. Building on this capability, the expansion of the parameters and tokens that Language Models use to turn inputs into outputs - which has grown by a factor of 1000 in the last 8 years - allows these models to both digest and use enormous amounts of contextual information when answering questions.

The ability to digest large amounts of data and produce useful, grounded (high quality) outputs is one dimension of the potential AI solution to this problem. The other two dimensions - cost and time - mirror the 'Iron Triangle' of project delivery. As has been explored the cost of this level of intelligence is dropping dramatically and is likely to continue to do so, making this capability ever more readily available.

Inherent Complexity and Decreasing Certainty

Whilst AI cannot change the nature of the complexity of construction projects or the fact that certainty must decrease further into the future, the emerging ability of generative AI models to reason in a sequential manner represent major progress. The most recent generation of reasoning models can break down complex tasks into their constituent parts, which enables specially trained generative AI models to produce schedules that are logically linked, reasonable in their chains of dependence (e.g. 'concrete pouring' must always come before 'concrete curing'), and inclusive of the required project scope.

In fact, the most recent generation of models have utilised a reinforcement learning approach to execute these more complex tasks. Reinforcement learning has previously been used to develop models capable of carrying out a schedule risk analysis [8]. The applicability of these models to the adjacent domain of schedule risk analysis is indicative of the potential value to planning of a well designed and rigorously tested planning tool that leverages similar models.

Advances in AI technology enable Language Models to deal with increasingly large volumes of data and to form complex relationships from this data. Combined, these capabilities allow AI systems to generate schedules of varying depth and detail, as shown by Schedule Studio. This means that these systems will also be able to generate plans in a rolling wave manner. Crucially, however, they will be able to explore an essentially infinite number of options that meet the contractual requirements by generating schedules iteratively. This ability may open the door for progressive proactive iteration of the schedule, only building out the detail when necessary and when future events are less at the whims of exogenous shocks.

Time Constraints & Strategic Planning

If AI systems can digest and understand large volumes of complex data and then apply logical and hierarchical reasoning to produce logical sequences that build up into a coherent construction project schedule then it is logical to ask how this technology can be used to not just augment existing workflows but to open up a new paradigm of construction project planning.

This paper previously noted that the inherent complexity of larger construction project schedules and the time constraints inherent in the reporting cadence of most projects made it almost impossible to routinely engage in strategic foresight and test delivery options. Instead, the baseline schedule is executed against until either the project ends (at which point the often adversarial delay claims process can take many more years) or delivery slips so far behind schedule that scope has to be removed via change order or a rebaselining exercise is carried out.

While the ability to rapidly generate and iterate on schedules will undoubtedly be useful in a re-baselining scenario, this technology will likely be utilized to iterate on the schedule on a monthly basis, exploring delivery options and being proactive rather than reactive. Given the work in the academic field of Decision Making Under Deep Uncertainty (DMDU) that has already been carried out to develop the Dynamic Adaptive Planning (DAP) framework - where plans are built to be flexible and to adapt to changing conditions rather than rigid and inflexible - this idea already has a sound intellectual footing.

Conclusion

The research outlined in this paper is the culmination of two years of full-time research into the underlying technologies that enable AI generated schedules, their applications, and their integration to develop the Schedule Studio platform - an end to end solution for using AI to develop construction project schedules from scratch using the source documentation.
The solution presented has demonstrated the feasibility of AI generated schedules and established the current best in class performance. The strengths and weaknesses of this technology have been explored and the likely technology developments that will enhance AI generated schedules in the near future.

Whilst the current technology is not sufficiently mature to be used on large construction projects today, the technology trends make it inevitable that in the near future AI schedule generation technology will reach parity with human expertise. This, however, is not a sufficient driver for widespread adoption, and hence development, of the technology. The rapidly decreasing cost of intelligence from these models will be the primary driver of further development. With thin margins for contractors and high risk for all stakeholders in a construction project, it is inevitable that once the technology underpinning schedule generation reaches maturity its low cost will be a driver for widespread adoption.

Future work in AI-generative scheduling could focus on several key areas. Integrating resource constraints into the scheduling process would enable the creation of more realistic and feasible schedules. This could involve developing a feedback loop between the scheduling model and LLMs, allowing for the identification and resolution of resource conflicts. Additionally, fine-tuning models on historical schedule data could enhance their ability to learn and replicate typical sequencing patterns, leading to improved schedule quality.

Further, leveraging historical data to generate data-driven duration predictions could increase schedule accuracy and proactively highlight potential conflicts with contractual milestones. Investigating methods for integrating company-specific experience into generative scheduling products would also enhance their applicability and effectiveness. Finally, developing techniques for capturing and presenting the assumptions made by AI scheduling models could improve transparency and facilitate collaboration between project teams and AI systems.

It is not foreseeable that AI will replace human planners on construction projects, in fact quite the opposite will be the case. The acute labour shortage in the global construction industry [7] will ensure that talented planners are elevated by AI, rather than replaced. As has been seen in the schedule risk assessment domain [10] the tasks executed by planners will move from reactive and process driven to proactive and creative explorations of delivery scenarios.

References

[1] PwC, Capital Project and Infrastructure Spending: Outlook to 2025, PwC, Report, 2014

[2] L. Woetzel, N. Garemo, J. Mischke, P. Kamra, and R. Palter, "Bridging Infrastructure Gaps: Has the World Made Progress?", McKinsey Global Institute, Report, Oct. 2017

[3] B. Flyvbjerg, N. Bruzelius, W.Rothengatter, "Megaprojects and Risk: An Anatomy of Ambition". 8th edn., Cambridge 2010

[4] Hollmann, J.: "Alternate Methods for Integrated Cost & Schedule Contingency Estimating". https://validest.com/uploads/1/3/6/0/136072948/hollmann_alternate_methods.pdf, last accessed 2024/03/12.

[5] Flyvbjerg, B. "Top Ten Behavioral Biases in Project Management: An Overview". Project Management Journal, 52(6), 531-546.

[6] Bertelsen, S. "Construction As A Complex System", in International Group for Lean Construction, Blacksburg VA, 2003

[7] CIOB, "The impact of the ageing population on the construction industry", CIOB, 2015

[8] Amratia, D., et al. "Practical Implementation of AI-based risk management processes on construction megaprojects". AACE International Technical Paper, Atlanta, 2024

[9] Hovhannisyan, V. et al. "Data-Driven Schedule Risk Forecasting for Construction Mega-Projects". AACE Conference & Expo, Chicago, 2023

[10] A. Mosca, R. Phillips, V. Hovhannisyan. "Quantitative Schedule Risk Analysis Using Artificial Intelligence Trained on Historical Data". Canadian Society of Civil Engineering Annual Conference, 2024

[11] Brown, Tom B., et al. "Language Models Are Few-Shot Learners". arXiv:2005.14165, arXiv, 2020.

[12] S. Laryea, "Quality of tender documents: case studies from the UK". Construction Management and Economics, 29 (3). pp. 275-286

[13] C. Smith, (2023). "Taylor Woodrow launches new energy and environment focused growth strategy". Available: https://www.newcivilengineer.com/latest/exclusive-taylor-woodrow-launches-new-energy-and-environment-focused-growth-strategy-02-03-2023/. Copy included in Appendix.

[14] Griffith A. F. "Scheduling practices and project success". In AACE International Transactions, 2005.

[15] D. Scott, "An investigation into project scheduling practice in the UK construction industry". MSt, Centre for Construction Engineering and Technology, Cambridge Univ., Cambridge, 2020

[16] W. E. Walker, V. A. W. J. Marchau and Jan H. Kwakkel, 'Dynamic Adaptive Planning (DAP)' in "Decision Making under Deep Uncertainty From Theory to Practice", 2019

[17] J, Olsson. (2020, September). "Lessons learned: improving the process". Construction Journal [Online]. Available: https://ww3.rics.org/uk/en/journals/construction-journal/lessons-learned--improving-the-process.html. Copy included in Appendix.

[18] Jaipuria, N. (2024). "The plummeting cost of intelligence". Available: https://www.wing.vc/content/plummeting-cost-ai-intelligence. Copy included in Appendix.

[19] OpenAI, et al. “GPT-4 Technical Report”. arXiv:2303.08774, arXiv, 2024.

[20] R. Anil, et al. “Gemini: A Family of Highly Capable Multimodal Models”. arXiv:2312.11805, arXiv, 17 June 2024

[21] Gao, Yunfan, et al. “Retrieval-Augmented Generation for Large Language Models”: A Survey. arXiv:2312.10997, arXiv, 2024.

[22] Huang et al. “LayoutLMv3: Pre-training for Document AI with Unified Text and Image Masking”, Proceedings of the 30th ACM International Conference on Multimedia, 2022.

[23] A. Vaswani, “Attention is all you need”, Advances in Neural Information Processing Systems, 2017.

[24] J. Wei et al., “Emergent abilities of large language models”, arXiv preprint arXiv:2206. 07682, 2022.

[25] H. Sajjad, N. Durrani, F. Dalvi, F. Alam, A. R. Khan, and J. Xu, “Analyzing encoded concepts in transformer language models”, arXiv preprint arXiv:2206. 13289, 2022.

[26] K. Park, Y. J. Choe, Y. Jiang, and V. Veitch, “The Geometry of Categorical and Hierarchical Concepts in Large Language Models”, arXiv preprint arXiv:2406. 01506, 2024.

[27] J. Devlin, “Bert: Pre-training of deep bidirectional transformers for language understanding”, arXiv preprint arXiv:1810. 04805, 2018.

[28] S. Minaee et al., “Large language models: A survey”, arXiv preprint arXiv:2402. 06196, 2024.

[29] J. Gerstmayr, P. Manzl, and M. Pieber, “Multibody Models Generated from Natural Language”, Multibody System Dynamics, vol. 62, pp. 249–271, 01 2024.

[30] L. Liu, X. Liu, J. Gao, W. Chen, and J. Han, “Understanding the difficulty of training transformers”, arXiv preprint arXiv:2004. 08249, 2020.

[31] V. Samborska and M. Roser, “Scaling up: how increasing inputs has made artificial intelligence more capable”, Our World in Data, 2024.

[32] P. Villalobos, A. Ho, J. Sevilla, T. Besiroglu, L. Heim, and M. Hobbhahn, “Will we run out of data? Limits of LLM scaling based on human-generated data”, arXiv preprint arXiv:2211. 04325, pp. 13–29, 2024.

[33] R. Fellows and A. M. M. Liu, “Managing organizational interfaces in engineering construction projects: addressing fragmentation and boundary issues across multiple interfaces”, Construction management and economics, vol. 30, no. 8, pp. 653–671, 2012.

[34] P. Kaur, G. S. Kashyap, A. Kumar, M. T. Nafis, S. Kumar, and V. Shokeen, “From Text to Transformation: A Comprehensive Review of Large Language Models’ Versatility”, arXiv preprint arXiv:2402. 16142, 2024.

[35] B. Peng, J. Quesnelle, H. Fan, and E. Shippole, “Yarn: Efficient context window extension of large language models”, arXiv preprint arXiv:2309. 00071, 2023.

[36] J. Ye et al., “A comprehensive capability analysis of gpt-3 and gpt-3.5 series models”, arXiv preprint arXiv:2303. 10420, 2023.

[37] G. Team et al., “Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context”, arXiv preprint arXiv:2403. 05530, 2024.

[38] J. Jiang, K. Zhou, Z. Dong, K. Ye, W. X. Zhao, and J.-R. Wen, "StructGPT: A general framework for large language models to reason over structured data", arXiv preprint arXiv:2305. 09645, 2023.

[39] H. Ko, H. Yang, S. Han, S. Kim, S. Lim, and R. Hormazabal, "Filling in the gaps: Llm-based structured data generation from semi-structured scientific data", in ICML 2024 AI for Science Workshop, 2024.

[40] J. Guo, L. Du, H. Liu, M. Zhou, X. He, and S. Han, "Gpt4graph: Can large language models understand graph structured data? an empirical evaluation and benchmarking", arXiv preprint arXiv:2305. 15066, 2023.

[41] Y. Sui, M. Zhou, M. Zhou, S. Han, and D. Zhang, "Table meets LLM: Can large language models understand structured table data? a benchmark and empirical study", in Proceedings of the 17th ACM International Conference on Web Search and Data Mining, 2024, pp. 645–654.

[42] A. Plaat, A. Wong, S. Verberne, J. Broekens, N. van Stein, and T. Back, "Reasoning with large language models, a survey", arXiv preprint arXiv:2407. 11511, 2024.

[43] J. Huang and K. C.-C. Chang, "Towards reasoning in large language models: A survey", arXiv preprint arXiv:2212. 10403, 2022.

[44] T. Zhong et al., "Evaluation of OpenAI O1: Opportunities and challenges of AGI", arXiv preprint arXiv:2409. 18486, 2024.

[45] A. Liu et al., "Deepseek-v3 technical report", arXiv preprint arXiv:2412. 19437, 2024.

[46] Shao, Zhihong, et al. "DeepSeekMath: Pushing the Limits of Mathematical Reasoning in Open Language Models". arXiv:2402.03300, arXiv, 2024.

Gerard Cardoso i Negrie - Head of Applied AI - nPlan
Damian Borowiec - Senior ML Research Engineer - nPlan
Rhys Phillips - Senior Client Manager - nPlan
Leonie Anna Mueck - VP Product - nPlan
Dev Amratia - CEO & Co-Founder - nPlan

Appendix

C. Smith, (2023). "Taylor Woodrow launches new energy and environment focused growth strategy"

LATEST

Exclusive | Taylor Woodrow launches new energy and environment focused growth strategy

02 MAR, 2023 | BY CLAIRE SMITH

Work in the energy sector is expected to account for 50% of workload at Taylor Woodrow within the next three years according to plans set out in a new strategy for growth unveiled by the firm today.

Taylor Woodrow managing director Phil Skegg told *NCE* that the business is less focused on achieving specific financial gains in that timeframe and more focused delivering sustainability in the mix of the work it undertakes.

"We must not focus on growth but on our contribution to the UK's energy transition," said Skegg.

This year the firm expects its turnover from energy related projects to be around 40%, which marks a significant change on 2017 where just 7% of its turnover was derived from the sector.

"We want to move over 50% of our activity into the energy market within the next three years in line with the UK's growing energy investments and stated long-term energy transition," explained Skegg. "There is no immediate focus on business growth; instead, it is about a repositioning in the markets that we operate."

Skegg said that the energy growth aims to build on the firm's strength in the rail, local and strategic highway infrastructure sectors.

Following on from launch of Taylor Woodrow's new environment strategy and framework at the start of the year, Skegg added that the new focus of the business will also have a bearing on future client selection to ensure there is "a strategic alignment of the business's ambitions regarding the environment".

The environmental strategy set out how Taylor Woodrow aims to find new ways of working that are better for the environment and to create links across the organisation to share learning on a continual basis. The firm said that this innovation and collaboration was key to avoiding "green washing" and work to deliver demonstrable incremental improvements at every stage of the infrastructure lifecycle.

To underline Taylor Woodrow's commitment, the firm has set out some clear net zero and environmental targets, which include:

- Reducing direct greenhouse gas emissions (scopes 1 and 2) by 40% by 2030 compared to 2018 levels
- Reducing indirect emissions (scope 3) by 20% by 2030 compared with 2019 levels
- By 2030 90% of all concrete used will be low carbon concrete
- Ban single use plastics from sites this year
- Reduce waste intensity year on year to achieve zero avoidable waste by 2040
- 100% of non-hazardous, construction, demolition and excavation Waste diverted from landfill by 2028
- Minimum of 10% Biodiversity net gain across all relevant projects by 2030

Supporting Skegg in delivering on the energy growth and net zero targets is a new leadership team, which also helps integrate the regional civils contracting businesses of Eurovia and SWH, with Taylor Woodrow, to create a stronger organisation. Taylor Woodrow has announced the appointments of Craig Prangley as operations director – projects, Louise Arrowsmith as operations director – regions, Jez Haskins in the role of business development director, Gareth Wagland is the pre-construction director and Millan Martin takes on the role of engineering director.

Skegg said: "The outcome will be a single team delivering diverse energy and public transportation projects, that enhance people's lives."

In making the announcement about the firm's new strategy, Skegg has also emphasised the importance of skills and industry research to support the strategy.

According to Skegg, recruitment, skills and training is fundamental to that plan. "Taylor Woodrow supports degree apprenticeships and graduate apprenticeships and has internal processes to ensure that skills and techniques such as slip forming are passed on," he said.

Alongside traditional engineering Taylor Woodrow will also be collaborating with its partners in the research and development of energy efficient materials and off site construction to minimise the environmental impact of its processes and help create a more sustainable supply chain.

Skegg added: "Wherever we can, we seek a sustainable solution for the engineering and design problems we have to overcome. That means being creative, it means being collaborative across our sector and beyond, and it means a commitment and a respect for the natural environment we work in as we create the built environment infrastructure to support the lives of people in the UK. It is about the right civil engineering solution. That's what Taylor Woodrow is about. And that's how, together as a team, Taylor Woodrow is forming a strong, sustainable business creating a positive legacy, I am very honoured to lead a strong senior leadership team, along with a very experienced business development, work winning and engineering teams, supporting all our project teams."

Like what you've read? To receive New Civil Engineer's daily and weekly newsletters click here.

J, Olsson. (2020, September). "Lessons learned: improving the process"

RICS

CONSTRUCTION JOURNAL

Lessons learned: improving the process

The lessons learned from projects are extremely valuable but how can they be best captured, disseminated and used to drive real improvement in the construction industry?

Author: Josh Olsson

16 September 2020

Learning and development | Project management | Projects and people

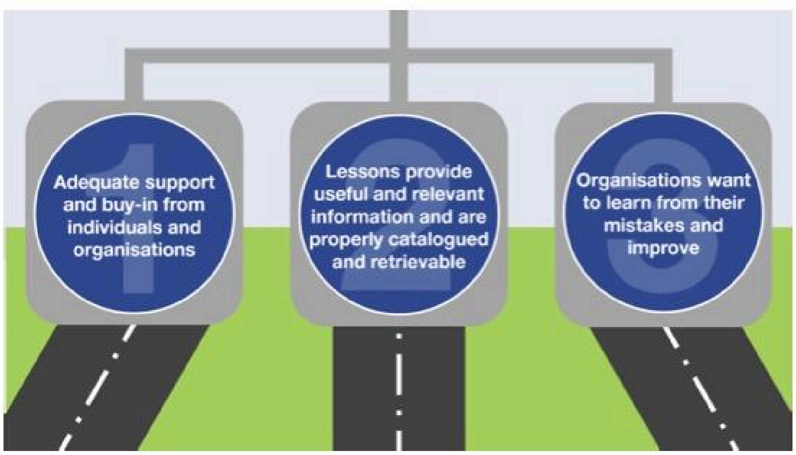
208

Phase 1 Report, for example, politicians of all parties repeatedly noted that lessons had to be learned from Grenfell to prevent it from happening again. However, it was also noted that learning the lessons would not be simple: one participant suggested that the government's ideological preconceptions and vested interests might skew the lessons learned process; another noted the fact that the lessons had clearly not been learned from the Lakanal House fire; and another pointed to the significant tension between allowing the time to complete a thorough review into Grenfell and the need to take immediate action to address existing safety issues in residential buildings.

A particularly interesting viewpoint was expressed by Emma Maier, editor-in-chief of Inside Housing: "It is clear that many of the lessons have not been learned … At times it has been difficult to get traction. The understandable focus on culpability and blame sends organisations into lockdown. It can get in the way of finding out what went wrong and sharing crucial details that could improve safety in the future... The Lakanal inquiry took 4 years, the court case took 8. The sad paradox is that the desire to gain understanding can make it harder to learn the lessons to prevent further tragedy."

The themes that emerge from Grenfell are a microcosm of the wider topic of lessons learned in general. While it is universally acknowledged that lessons learned are of intrinsic value and that a lessons learned workshop should be a key part of virtually any project, it is less clear how a project manager is to ensure that lessons are best captured, disseminated, and – most importantly – actually learned from and used to drive real improvement.

I believe there are 3 checkpoints that a specific lesson has to pass through in order to ensure that it can be effectively learned and used.



Thi[illegible]oint is right at the beginning of the lessons learned journey, and the requirement for passage through it is simply that there is sufficient buy-in to the very concept of learning lessons from projects.

While this might appear to be a relatively low threshold for entry, it is clear that many projects are not critically assessed with the objective of understanding and capturing the most important lessons. The RICS *Lessons Learned* guidance note divides the reasons for this into individual and organisational factors.

Individual factors include:

- project stakeholders considering the process a waste of time
- being fearful of being punished for admitting to a mistake
- being more interested in self-promotion in front of the client than a deep dive into project failings.

Organisational factors, on the other hand, include:

- having a culture that is not interested in learning
- having a culture that is characterised by short-termism and a focus on either conflict avoidance or finger pointing
- not having the resources to fully support the lessons learned process.

Since projects are often thought of as unique endeavours, there can be a tendency to get lost in their exceptionalism and believe that there is little which can be applied outside of the specific context of the project. This viewpoint can be amplified if an organisation is participating in a project that is considered a one-off. What is the point of doing a thorough lessons learned review for a small bridge project procured under an NEC form of contract, for example, when the vast majority of a firm's work is fit-out work procured under JCT?

Often, shortly after contractual completion most members of a project team are immediately redeployed on to their next assignment. It can be extremely difficult to get either attendance or sustained engagement in the lessons learned process from individuals who are required to hit the ground running on their new project.

"Failure to pass through the first checkpoint means that either lessons are not captured in the first place or, if they are, the process takes place in an environment which will not produce optimal results"

Unfortunately, failure to pass through the first checkpoint means that either lessons are not captured in the first place or, if they are, the process takes place in an environment which will not produce optimal results. There is a serious risk that the lessons learned workshop might be rushed, poorly attended, and undertaken in the manner of a blame game, rather than a collaborative and comprehensive effort to genuinely understand how to improve performance.

However, there are practical things which project managers can do to ensure that the lessons learned process has the best chance of obtaining useful outputs.

RICS

lifecycle. Consultants' appointments and contractors' prelims should stipulate engagement in specific lessons learned activities, and they should also be clearly referenced in both the programme and in the project execution plan. Moreover, whatever the level of formal buy-in to the lessons learned process, the project manager will also need to use all of their stakeholder management, communication and conflict avoidance skills to ensure that the right people meaningfully participate in lessons learned activities.

The project manager should also look to best practice guidance for conducting a lessons learned session. For instance, the RICS' guidance note provides common sense instructions for convening a workshop, including clearly communicating the expectations and benefits to potential participants, collecting a list of issues to help create a thematic framework, setting clear rules of engagement on the day, and the use of an independent facilitator where possible. This guidance note also provides some sample slides which outline what the workshop is and is not, its rules and the expected inputs and outputs.

208
Shares

RICS

Once a lesson has been comprehensively discussed by the relevant project team members, the next checkpoint is to ensure that it is properly documented and stored so that it can be easily accessed when it is relevant to a future project.

In an article for the Project Management Institute, Sandra F. Rowe and Sharon Sikes suggested the use of several techniques to help achieve this.

These techniques include:

- project survey forms with standardised categories to structure lessons learned reports
- lessons learned templates which use consistent fields such as category, lesson learned, action taken, reasoning for action taken, root cause and key words
- the use of a dedicated lessons learned repository at an organisational level.

Rowe and Sikes stressed the importance of key words in particular, as these are used to ensure that relevant lessons can be found when future project teams search for relevant terms.

This focus on a systematic approach to cataloguing lessons learning in order to facilitate accessibility and retrievability is also echoed by the Association for Project Management. In its Body of Knowledge (6th edition), it recommends the following approach for dealing with lessons learned.

"The key steps involved in ensuring that knowledge is captured and shared include:

- establishing ownership of knowledge management and the systems that support it;
- implementing mechanisms for finding external knowledge and making it relevant internally;
- identification and extraction of key lessons from projects, programmes and portfolios, including contextual information;
- structuring and storing knowledge so that it can be accessed easily;
- maintenance of the knowledge repository to ensure it is up to date;
- embedding processes that ensure knowledge is used effectively."

While it is clearly common sense that lessons should ideally be captured in a way that enhances their usefulness and stored in a way that maximises their retrievability, the reality appears to be quite far removed from this ideal.

Martin Paver, a specialist in data-centric project management, has pulled together a database of more than 10,000 lessons learned from a variety of sources, including National Audit Office reports, parliamentary reports, and Freedom of Information requests. Paver's review of his database has led him to conclude that many of the lessons lack forensic insight, context and root cause analysis, that they are captured without using a standardised format, and that the focus is on process rather than outcomes.

"While it is clearly common sense that lessons should ideally be captured in a way that enhances their usefulness and stored in a way that maximises their retrievability, the reality appears to be quite far removed from this ideal"

sigr RICS iers of BBC employees and third parties… It is estimated that completing these tasks could take 26 days to retrieve and compile the requested information".

If it takes a major organisation 26 days to retrieve the lessons learned for recent and important projects, then the question can legitimately be asked of how these lessons are actually able to be learned. And this does not even factor in 2 other potential barriers.

The first of these is that – from a consultancy perspective at least – many clients insist on NDAs and confidentiality agreements, and this, along with the impact of GDPR, can have an effect on the quality of lessons learned, or at least make capturing a good lesson require significantly more effort. Some lessons learned reports may have to be cleansed of specific client details or personally identifiable information and have the contextual details removed in order to comply with framework agreements, service instructions and current legislation.

Second, even assuming that the lessons are properly categorised, tagged and stored in a repository where they can be easily accessed, the usefulness of the lessons are still highly dependent on their retrievability. For instance, if the lessons are simply saved in a large Excel sheet or as separate documents on a SharePoint site, then there is a very real danger that searching for common terms or keywords could – depending on the size of the database – bring up dozens or even hundreds of largely irrelevant results. The search functionality for these kinds of databases can be slow, non-intuitive and unable to prioritise the results to provide the requestor with the most useful lessons.

This was a very real issue experienced by NASA, which has a lessons learned database of almost 10m documents dating back over the past 60 years. The database historically used a PageRank algorithm which prioritised search results according to how frequently they had been accessed, and this meant that for one specific query a NASA engineer had to read through around 1,000 individual documents to see if they contained any relevant information.

NASA solved this problem by investing in new software to better visualise the correlations between various lessons learned in order to improve the relevance of the results. This drives home the point that it is very important for all construction industry organisations to evaluate their own lessons learned databases to ensure that the opportunity to retrieve relevant lessons is maximised. There are numerous ways of doing this – and a plethora of different digital applications that could be used – but the central objective has to be to avoid vital information being siloed away and forgotten.

Another key action that can be undertaken at an organisational level is to ensure that standardised tools and templates are used for the lessons learned process to ensure data consistency. Lessons learned resources should be included within project and programme management toolkits, and a key requirement for colleagues should be regular and accurate submission of lessons learned into a centralised repository.

At an individual level, the main requirement of the project manager is to ensure both that a lessons learned process is properly adhered to and that the lessons learned themselves contain enough

Ass RICS n organisation has embedded a mature lessons learned process and maintains an accessible knowledge platform, and that individuals regularly and committedly participate in lessons learned reviews, then there is only one checkpoint left in the journey to ensuring that the most important lessons are actually learned and used. In many ways, however, this checkpoint is the most vital one of all. To gain passage through this checkpoint, organisations need to view the whole lessons learned process not just as a tick-box best practice exercise, but as an essential way of learning from their mistakes and achieving continuous improvement. They must want to learn from their lessons and to implement any required changes.

An interesting insight into the reluctance of many organisations to make the changes recommended by their lessons learned process is provided by Kevin Pollock's report on interoperability in the emergency services. Pollock ultimately concluded that: "The consistency with which the same or similar issues have been raised by each of the inquiries is a cause for concern. It suggests that lessons identified from the events are not being learned to the extent that there is sufficient change in both policy and practice to prevent their repetition.

"[In the majority of lessons learned reviews] the doctrine and prescriptions are often structurally focused, proposing new procedures and systems. But the challenge is to ensure that in addition to the policy and procedures changing, there is a change in organisational culture and personal practices. Such changes in attitudes, values, beliefs and behaviours are more difficult to achieve and take longer to embed. However, failure to do so will result in the gathering of the same lessons which repeat past findings rather than identifying new issues to address and continuously improve the response framework."

A 2018-19 audit into the London Fire Brigade found that: "The London Fire Brigade has clearly learned lessons from the Grenfell Tower incident. However, it has been slow to implement the changes needed, which is typical of the brigade's approach to organisational change." In this instance the organisation has the capability and resources to learn the lessons from Grenfell, but perhaps did not, at the time, have the required level of commitment to force through what might be complex, far-reaching and difficult changes.

"Organisations need to view the whole lessons learned process not just as a tick-box best practice exercise, but as an essential way of learning from their mistakes and achieving continuous improvement"

Part of the solution for this might be for organisations to have dedicated resources for evaluating lessons learned, reporting trends and patterns, and ultimately either implementing the changes themselves or offering clear recommendations to those with decision-making authority. As Rowe and Sikes suggest: "... organisations need to dedicate a resource or resource(s) to begin the analysis of documented lessons learned. The purpose of the analysis is to identify actions that can be taken within the organisation to strengthen weak areas of knowledge and implementation during each project. This can be done through enhanced training of project managers and/or team members; this includes project sponsors and champions. It may mean added or improved procedures and processes.

RICS

Firms will also need to transform themselves into what Pollock terms a "learning organisation". This type of organisation will "ensure that the lessons learned will result in changes to the organisational culture, norms and operating practices. These will be successfully embedded in the values and beliefs of the organisation and those who work in it."

At an individual level, project managers can also have an impact in terms of fostering a culture receptive to learning from past experience and committed to continuous improvement. For instance, Paver recommends that positive lessons in particular should be shared widely within an organisation through the use of technical papers, lunch and learn sessions, campaigns and training, and by developing wiki content and videos. All of this can help nudge an organisation in the right direction.

My experience of lessons learned

My own experience of the lessons learned process is varied. In some instances – and particularly when delivering successive projects on a framework – I have led comprehensive and collaborative workshops, captured important and relevant lessons, and been able to take these learnings into the next scheme to demonstrably improve project delivery. On occasion, however, and particularly when undertaking one-off projects, I have found that it has been very difficult to get stakeholders to buy into the lessons learned process, that the workshops have been both poorly attended and full of recriminations – particularly when something has gone wrong – and that the lessons learned report is issued and promptly forgotten about.

There are many reasons for this, but the stakes are so high in the construction industry that it is extremely unfortunate that the hard-won experience and insight accumulated by all construction professionals cannot be better leveraged to learn from our experiences and to improve quality, safety and efficiency across the industry.

Facilitating the journey of a lesson through the 3 proposed checkpoints is by no means an easy task, but it is a necessary one if construction industry professionals and organisations are to maximise their quality, safety and efficiency. The foregoing discussion outlines some of the potential barriers and possible solutions for ensuring that lessons are better learned and used, but this is just the tip of the iceberg in what is clearly a complex and multifaceted issue.

A useful next step would be an industry-wide review of the lessons learned from trying to implement a lessons learned culture, but in the interim I would be more than happy with a thoughtful debate on how to improve the effectiveness of the lessons learned process in the construction sector.

208
Shares

Jaipuria, N. (2024). “The plummeting cost of intelligence”

The plummeting cost of intelligence

Tanay Jaipuria
Author

NOVEMBER 4, 2024

THE PERCEPTION THAT AI IS EXPENSIVE IS RAPIDLY BECOMING OUTDATED. THROUGH A COMBINATION OF TECHNOLOGICAL BREAKTHROUGHS AND MARKET FORCES, WE'RE ENTERING AN ERA WHERE INTELLIGENCE COULD BECOME NEARLY FREE TO USE.

1. Competition and market forces

2. Improving efficiency in compute

3. The rise of smaller, smarter models

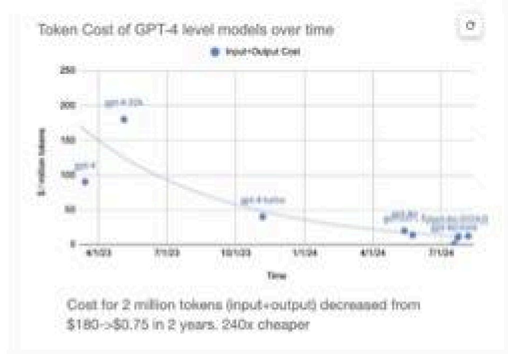
4. New model architectures

While foundation models are often perceived as costly and inference bills are rising, the reality is that the cost of intelligence is in free fall — and this trend shows no signs of slowing. Angel investor Elad Gil shared this chart on the cost of GPT-4 equivalent intelligence from Open AI, which has fallen 240x in the last 18 months, from $180 per million tokens to less than $1.

driving efficiency

5. On-device inference

Look beyond today's intelligence costs



Source: Elad Gil on X

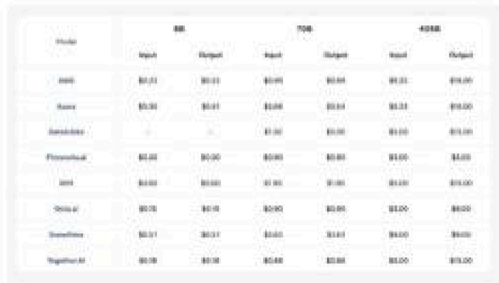
Several factors are driving the continued decline in the cost of AI intelligence. If this trend continues, we may reach a point where the cost of using AI models becomes negligible for most use cases, effectively approaching zero.

In this piece, we'll explore the key factors contributing to the declining cost of intelligence.

1. Competition and market forces

Competition has driven down costs as more companies enter the market, with open-source models from Meta and others now matching GPT-4's performance. This competitive pressure exists between model developers and the inference providers who run these models.

For example, the rate sheet below shows the variety of provider options for Llama 3.1.



Model	8B		70B		405B	
	Input	Output	Input	Output	Input	Output
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	-	-	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Source: Llama website

Similarly, for a given level of model quality, users have multiple options, which results in pricing pressure and some degree of convergence in prices, as in the chart below.

Source: Artificial Analysis

Even with OpenAI's recent release of o1, historical patterns suggest other models will eventually catch up. While costs may temporarily increase due to the model's higher computational demands at inference time, prices should ultimately continue their downward trend.

2. Improving efficiency in compute

Obviously, market forces alone aren't enough to lower prices. Optimizations at both the hardware and infrastructure layers have reduced the cost of running inference, allowing companies to pass these savings on to customers.

Significant hardware innovation is driving costs down, with specialized chips/ASICs like Amazon's Inferentia and new players like Groq. While these solutions are still emerging, they're already demonstrating dramatic improvements in both price and speed.

Amazon says their Inferentia instances deliver up to 2.3x higher throughput and up to 70% lower cost per inference than comparable Amazon EC2 options.

Similarly, as inference workloads are starting to scale up, and more talent is building in AI, we're getting better at utilizing GPUs more effectively, and getting more economies of scale and lower inference costs through optimizations at the software layer as well.

3. The rise of smaller, smarter models

Another key reason for the declining cost of AI is the improvement in performance for a given level of model size — and smaller models are getting smarter over time.

Here's one example: Meta's Llama 3 8B model essentially performs on par (or better than) their Llama 2 70B model, which was released a year earlier. So within a year, we got a model nearly one-tenth the parameter size that had the same performance. Techniques like distillation (using the outputs of larger models to fine-tune smaller, task-specific models) and quantization are making it possible to create increasingly capable compact models.

Notably, Llama 3.1 405B's license permits using its outputs to fine-tune other AI models, which further enables this trend.

According to Llama's community license agreement, "If you use the Llama Materials or any outputs or results of the Llama Materials to create, train, fine tune, or otherwise improve an AI model, which is distributed or made available, you shall also include 'Llama' at the beginning of any such AI model name.

4. New model architectures driving efficiency

There's also a push toward entirely new model architectures that promise to make AI even more efficient. While transformer models still dominate, new architectures such as state space models are emerging as strong contenders, with companies such as Cartesia leading the way.

These new approaches can be faster and more efficient, and they require less compute power to achieve comparable performance. As companies make more progress on these approaches, they could enable small, highly efficient and robust models that further reduce the cost of intelligence through lower inference costs.

As an example, some of the Mamba class of models in the 1.5B-3B parameter range outperform the leading transformer-based models of that size.

5. On-device inference

The future of AI isn't just about cloud-based models — it's increasingly about running AI directly on end-user devices. Apple has already announced two proprietary 3B parameter models that will run locally as part of their Apple Intelligence launch: one language model and one diffusion model (more details). While these models are currently limited to Siri and other first-party apps, Apple will likely open them to developers in the future.

Apple's benchmarks show that while on-device models can't be used for all queries, users tend to prefer them for many prompts, even compared to larger models, as below.

Source: Apple

In addition, as the chips on laptops and phones continue to get more powerful, coupled with models getting smaller and smarter as discussed above, a larger fraction of the most commonly used needs for intelligence can be handled with models running locally.

For example, companies like Ollama now enable users to run popular models, including Llama 3-8B, directly on their laptops.

Beyond offering users privacy and reduced latency, local processing will dramatically reduce costs. When AI runs directly on our devices, the cost of intelligence will effectively drop to zero.

Look beyond today's intelligence costs

As the price of AI continues to drop, we'll see a wave of new applications and industries embracing these technologies. My advice to founders and builders is not to focus too much on inference costs (as long as they aren't causing significant cash burn) and to avoid over-optimization too early, as these costs are dropping rapidly.

Instead, I encourage founders and builders to think about what use cases or additional features don't seem feasible yet, but are potentially unlocked as the cost of intelligence drops to one-tenth or one-hundredth of the current price, since we'll likely get there sooner than most people think.

If you're building a company that helps drive down these inference costs across any layer or leverages the future of low-cost intelligence to solve problems for end users, I'd love to hear from you - reach me on X.