A critique of AI tool errors in Critical Path Analysis theory – a project control case study [1]

Dr. Dimitris Antoniadis [2]

Abstract

As the development and implementation of Artificial Intelligence (AI) tools globally continues to rise, we are beginning to see trends in published research considering its benefits and drawbacks. Several bodies have developed rules underpinning its use in their environments, while others (including governments) are continuing to work on building further regulations. The author has reflected on the latest trends in this space, especially the use of AI by project management professionals. In the latter part of January 2024, the author embarked on an extensive review of AI's current abilities in delivering Critical Path Analysis (CPA) outputs by testing ChatGPT (v3.5), and in cooperation with others, on ChatGPT (v4), as well as several other AI project management AI tools. This article will highlight the limitations in ChatGPT's and other AI tools functionality in compiling CPA outputs. Using an example from Project Management Institute (PMI), the author will present these errors as well as a critique of the outputs.

Key words: AI, Artificial Intelligence, ChatGPT, Project Control, Critical Path Analysis (CPA)

Introduction and the Case Study

The aim and objectives of this article are, through a case study, to conduct a critique on AI tool outputs on the specific scheduling technique of Critical Path Analysis (CPA). As part of wider research interest in AI and its use in project management and project control the author was reviewing a series of slides issued by the Project Management Institute (PMI) with the title '***Generative AI Overview for Project Managers - Resources***' (PMI, 2023). These demonstrate the use of ChatGPT in project management, discussing the use of GenAI and generally DOs and DON'Ts in the use of GenAI.

In slide 12 of the PMI presentation a prompt is given (see Table 1 below) for ChatGPT to provide a Critical Path Analysis output for a hypothetical simple schedule of five (5)

[1] How to cite this paper: Antoniadis, D. N. (2024). A critique of AI tool errors in Critical Path Analysis theory – a project control case study; case study, *PM World Journal*, Vol. XIII, Issue VI, June.

[2] Director, DANTON PROGM Ltd, London, UK; Course Leader & Senior Lecturer, London Metropolitan College, London, UK; Correspondence: dnanton00@gmail.com; Website: www.danton-progm.co.uk

activities. The author will use the ChatGPT output provided in the PMI slides (see Figure 2) to highlight an interesting 'error' in the resulting outputs.

I am a project manager for a construction development project. The project has five activities: • Activity 1 {predecessor = start and duration = 3 days} • Activity 2 {predecessor = Activity 1 and duration = 3 days} • Activity 3 {predecessor = Activity 1 and duration = 4 days} • Activity 4 {predecessor = Activity 2 and duration = 8 days} • Activity 5 {predecessor = Activity 3, 4 and duration = 4 days} Show me, in a table format, the dependencies between the tasks and their corresponding early start (ES), early finish (EF), late start (LS), late finish (LF) times; float for each task; and all the paths with duration. Solve this and show your work concisely with the critical path and all other paths in days. Highlight the critical path with the shortest duration and least float value of 0 in bold.

Table 1. The ChatGPT prompt to generate a simple five activity schedule (PMI, 2023).

The ChatGPT output, as presented by PMI, is shown below in two parts Figure 1 and Figure 2 for ease of reading.

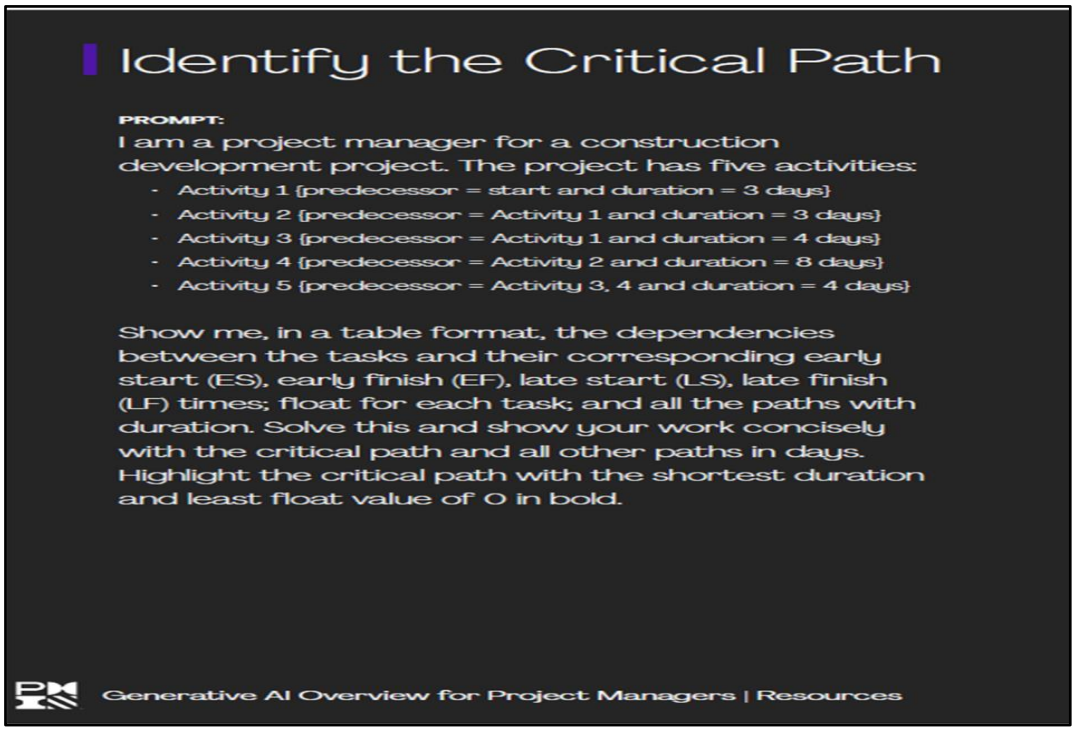
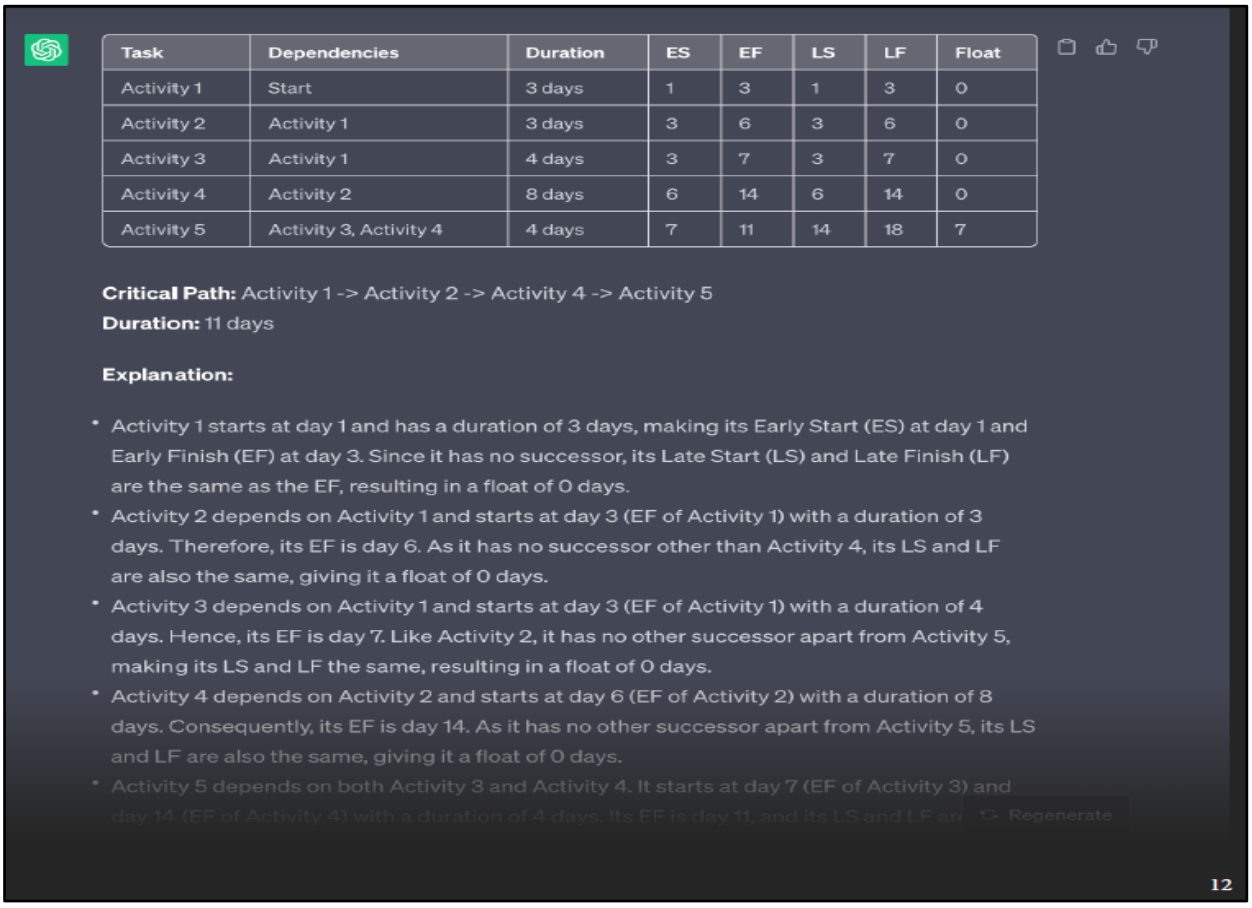


Figure 1. Part 1 of PMI Slide 12 with a ChatGPT query to generate the CPA of a five-activity network. (PMI, 2023)



Task	Dependencies	Duration	ES	EF	LS	LF	Float
Activity 1	Start	3 days	1	3	1	3	0
Activity 2	Activity 1	3 days	3	6	3	6	0
Activity 3	Activity 1	4 days	3	7	3	7	0
Activity 4	Activity 2	8 days	6	14	6	14	0
Activity 5	Activity 3, Activity 4	4 days	7	11	14	18	7

Figure 2. Part 2 of PMI Slide 12 with ChatGPT output for the CPA of five activity schedule (PMI, 2023)

Reflections on the Outputs presented

The author reviewed the output presented, and several basic errors were immediately obvious. Figure 3 presents the points in question, which are discussed in more detail below.

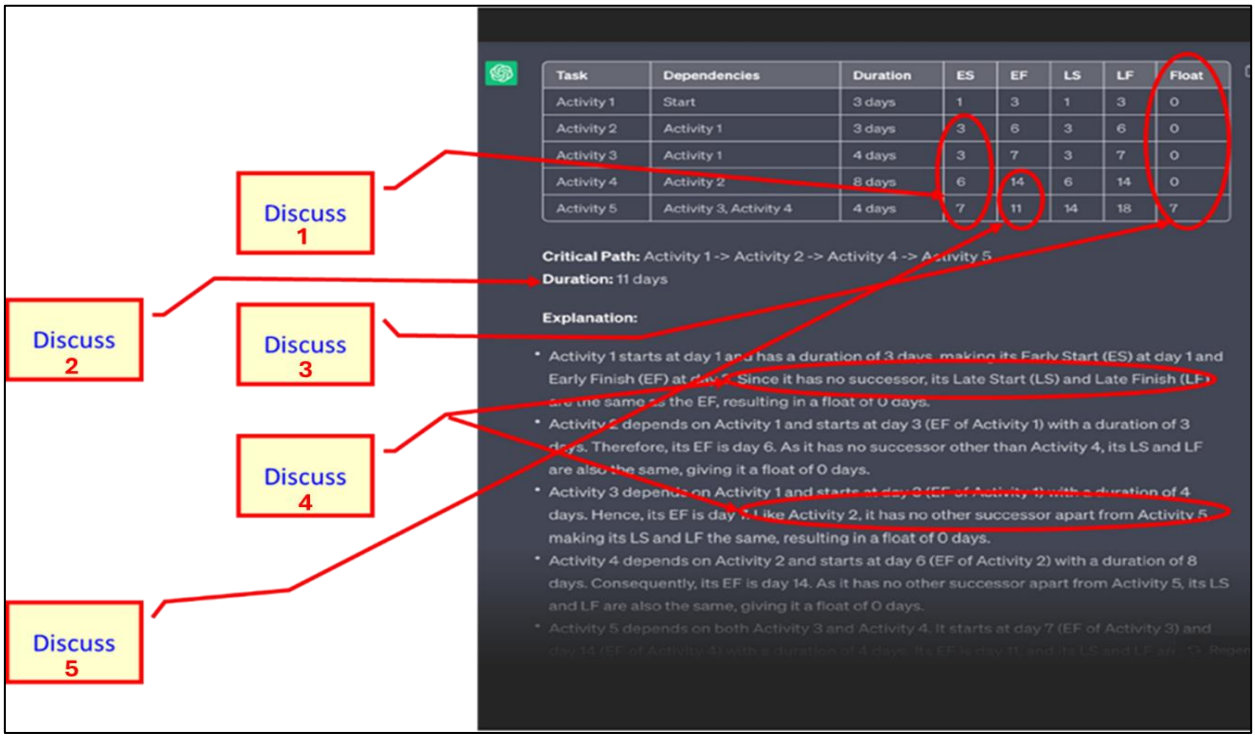


Figure 3. Author's identification of errors in the output.

Some of the questions raised by the AI tool output are shown as discussion points 1 to 5:

1. Discussion point 1: How is it possible for ES (Early Start) dates of activities to be the same with the EF (Early Finish) dates of the preceding activity?
   This practically and theoretically is an error.

2. Discussion point 2: How can the total duration of the schedule be 11 days when, the EF of activity 4 is 14 days and the total duration of the critical path is 18 days?
   This is a serious unchecked error,

3. Discussion point 3: How is it possible for the last activity in a schedule to have Total Float? This is a theoretical error, again unchecked.

4. Discussion point 4: Under the 'Explanations', where is the logic in a) the claims that Activities 1 through 4 have no successors and b) that as a result the LS and LFs are the same? What does this mean? All activities listed, apart from the final Activity 5, have successors, therefore the ChatGPT statement is wrong.

5. Discussion point 5: How is it possible for Activity 4 EF to be greater than the EF of the final activity (Activity 5)? This is another theoretical error as Activity 5 is the final activity and therefore the EF has to be later/greater than the EF of Activity 4.

The author considered that there must be an error in the PMI slides presented and decided to test the same prompt in ChatGPT (v3.5) during the latter part of January 2024.

The test was conducted a number of times repeatedly and despite using the same prompt, ChatGPT's (3.5) response changed each time and each response continued to present inaccuracies / wrong results.

The author considered that the results needed to be cross checked and two other individuals were asked to conduct the same test (in the same period – latter part of January 2024). Using the same prompt and in addition to testing ChatGPT (v3.5) they were free to check the outputs from any other AI tool they had available in order to achieve a wider perspective on the AI tool performance.

The two individuals in addition to ChatGPT (v3.5) used three different AI tools:

- Test 1 - Project Management GPT,
- Test 2 - GPT 4, and
- Test 3 - a Project Management 'bot'.

Further to the above and in order to add to the wider checking the author conducted a test – Test 4 – using the PMI AI tool 'aiassistant' (which can be accessed in: https://aiassistant.pmi.org/chat).

The results from all tests are included below, and the author has introduced some markers to highlight the resulting errors.

Output from Test 1 – the Project Management GPT

Activity	Duration	ES	EF	LS	LF	TF
1	3	0 ??	2	0	2	0
2	3	3	5	3	5	0
3	4	3	6	10	13	7
4	8	6	13	6	13	0
5	4	14	17	14	17	0

The critical path is the sequence of activities that have zero total float. Here, it's Activities 1 → 2 → 4 → 5, with a total duration of 18 days (bolded in the table).

Figure 4. Output from Test 1 – Project Management GPT
(Date: latter part of January 2024).

The results from Test 1 – the Project Management GPT's provided the most accurate output from the tests carried out; however, the graphic shows the following errors were identified:

- The ES of Activity 1 is shown as zero (0) and
- Completion is shown as day 17 (see EF and LF of Activity 5) as opposed to day 18 (as per the AI tool worded comment shown at the bottom of Figure 4).

Further discussion of these results will be presented later.

Output from Test 2 – the classic ChatGPT (v4)

Activity	Predecessor	Duration (days)	ES	EF	LS	LF	Float
1	Start	3	0	3	0	3	0
2	1	3	3	6	3	6	0
3	1	4	3	7	5	9	2
4	2	8	6	14	6	14	0
5	3, 4	4	14	18	14	18	0

Paths and Their Durations:

- Path 1: Activity 1 -> Activity 2 -> Activity 4 -> Activity 5 = 3 + 3 + 8 + 4 = 18 days
- Path 2: Activity 1 -> Activity 3 -> Activity 5 = 3 + 4 + 4 = 11 days (Not continuous)

Critical Path (in bold):

- **Activity 1 -> Activity 2 -> Activity 4 -> Activity 5 (18 days)**

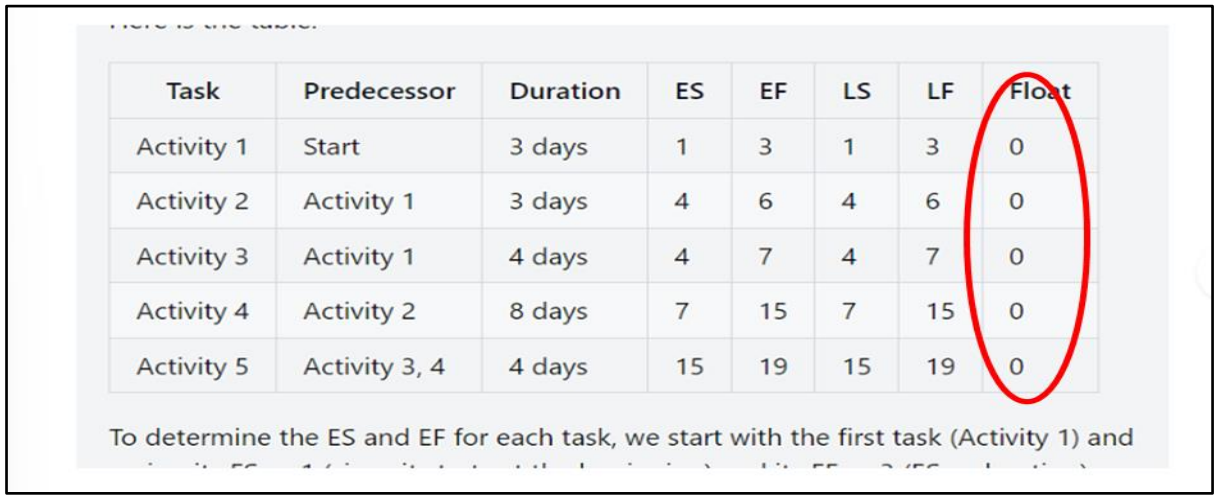
Figure 5. Output from Test 2 – the classic ChatGPT (v4)
(Date: latter part of January 2024).

The author has highlighted a number of errors numbered above as discussion points 1, 2 & 3:

- Discussion point 1: A completely incorrect Total Float figure is given against Activity 3. This should be 7 as per the output from Figure 4.

- Discussion point 2: All the standard dates (ES, EF, etc., as shown in Figure 5) were incorrect, with ES starting on the same day as the EF of the preceding activity. All ES dates should be commencing on the day after, however, further points on this will be addressed under Discussion.
- Discussion point 3: For all activities, the '*Duration (days)*' column is not accurately reflected in the ES and EF days. For example, Activity 2 has a 3-day duration. The ES Day is 'Day 3' and the EF day is 'Day 6, which means a 4-day Duration.
- Discussion point 4: The start of the schedule is shown as 'Day zero (0) and Completion is 'Day 18', which means a 19-day schedule, however the Total Duration is 18 days (see the Critical Path line).

Output from Test 3 – a well-known BOT



Task	Predecessor	Duration	ES	EF	LS	LF	Float
Activity 1	Start	3 days	1	3	1	3	0
Activity 2	Activity 1	3 days	4	6	4	6	0
Activity 3	Activity 1	4 days	4	7	4	7	0
Activity 4	Activity 2	8 days	7	15	7	15	0
Activity 5	Activity 3, 4	4 days	15	19	15	19	0

To determine the ES and EF for each task, we start with the first task (Activity 1) and

Figure 6. Output from Test 3 – a well know BOT (Date: latter part of January 2024).

The author has highlighted several areas and considerations below as discussion points 1 to 5:

1) Discussion point 1: The output shows no Total Float, which theoretically is inaccurate for this particular schedule and with the logic links stated in the prompt.
2) Discussion point 2: Activity 3 should have 7 days Float (as per output in Figure 4 and Figure 8 below), however this is not shown.
3) Discussion point 3: The ES of Activities 2 and 3 were correct. However, the ES of Activity 5 is incorrect, starting at the same time as the EF of Activity 4.

4) Discussion point 4: The EF of Activity 4 is incorrect. If the activity has an 8-day duration, the EF should be 14.

5) Discussion point 5: This same issue repeats with Activity 5. Based on a 4-day duration, the EF should be 18.

Outputs from Test 4 – the PMI AI tool 'aiassistant'

Task	Predecessor	Duration	ES	EF	LS	LF	Float
1	Start	3	1	3	1	3	0
2	1	3	3	6	3	6	0
3	1	4	3	7	5	9	2 ??
4	2	8	6	14	6	14	0
5	3, 4	4	14	18	15	19	1 ??

1

2

To find the critical path, we need to identify the longest path in the network. In this case, the critical path is:
Start - Task 1 - Task 2 - Task 4 - Task 5

3 The duration of this path is 14 days. All other paths have a shorter duration. The float value for each task is calculated by subtracting the EF of the current task from the ES of the next task.

Figure 7.1 Output from Test 4 – the PMI 'aiassistant' tool
(Date: latter part of January 2024).

The author has identified the following errors in Figure 7.1:

- Discussion point 1: The calculations of ES and EF are inconsistent, based on the duration of activities provided. For example, Activity 2, with 3 days duration should have an EF of 5. As it is the duration should be 4 days.
- Discussion point 2: More concerning is the Total Float calculations, particularly given the tool sets out flawed logic in how this should be calculated:
    - Total Float of one (1) day on the end Activity 5 is theoretically wrong. Activity 5 has no successor therefore, this should be 0, and Total Float on Activity 3 again is incorrect. It is the only activity that is not in the critical path and it should have Total Float of 7 days.
- Discussion point 3: In the worded response (see yellow highlighted sentence), the AI tool indicates that the critical path is 14 days long, again incorrect.

Despite the errors discussed above and as can be seen in Figure 7.2, the tool attempts to validate its calculations by providing relevant sources.

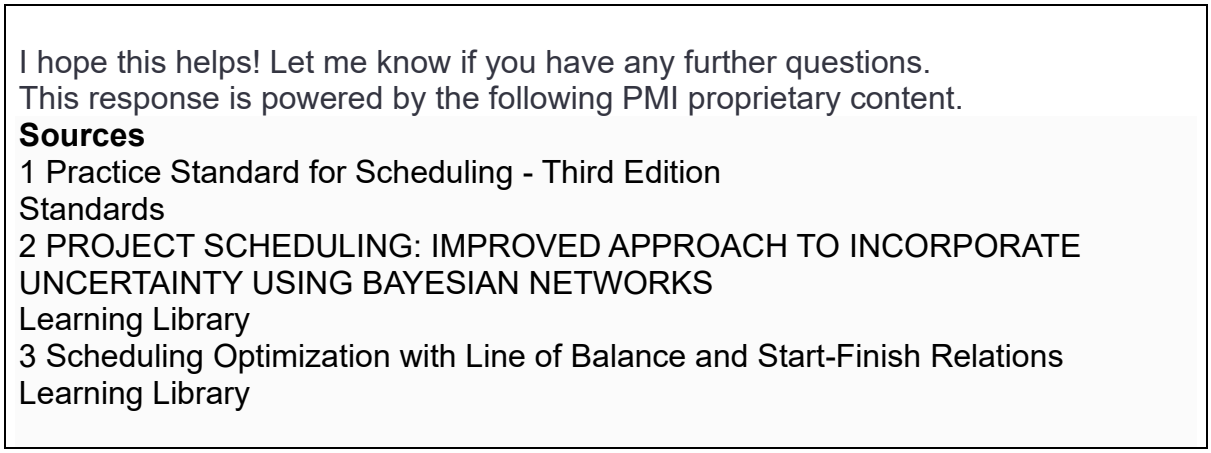

I hope this helps! Let me know if you have any further questions.
This response is powered by the following PMI proprietary content.
Sources
1 Practice Standard for Scheduling - Third Edition
Standards
2 PROJECT SCHEDULING: IMPROVED APPROACH TO INCORPORATE UNCERTAINTY USING BAYESIAN NETWORKS
Learning Library
3 Scheduling Optimization with Line of Balance and Start-Finish Relations
Learning Library

Figure 7.2 Output from Test 4 – the PMI 'aiassistant' tool indicating sources used for the result.

The author has checked reference 2 – '*Project scheduling: improved approach to incorporate uncertainty using Bayesian networks*' authored by Khodakarami, et al. (2007) – and this article has no relevance to the output the tool has provided. Was this a hallucination by the AI tool?

Further points relevant to practical scheduling and CPA

In addition to the above errors, there is also another consideration.

Practitioners and project management professionals know that the software tools that perform CPA use a calendar and, in the majority of cases, these are set up to have a working week of five days and a weekend. Therefore, it is known that activity durations should represent working days, and the scheduling calendar takes into consideration the weekend days off (Saturday and Sunday or Friday and Saturday). However, based on the outputs displayed above from both ChatGPT (Figures 4, 5 and 7.1) and the 'Bot' (Figure 6) do not recognise this.

Whatever happens in the Large Language Model (LLM) Machine Learning (ML) model(s) used, there is something fundamentally wrong with the processing of prompts and CPA calculations produce the wrong outputs.

Unfortunately, at this point in time, this is not a convincing case to encourage professionals to use the AI technology promoted in whatever form. Given the number of errors across the board, it is not just a case of asking professionals to use the technology with caution but ensuring that it delivers the required output(s) correctly.

In the past, similar mistakes have occurred with new technology being promoted as the answer to various issues too early. The author can see history repeating itself and practitioners ignoring and / or rejecting the latest technologies or perhaps worse, using the technology and obtaining the wrong results. A question here on this point; with the current technology what would have happened if the sample schedule was for 30 activities?

A correct CPA output from a simple scheduling software tool

To provide a more accurate response to the prompts as a comparison, the author modelled the example in the scheduling tool MSProject, and the output can be seen in Figure 8 below in a 'straightforward' logically linked barchart view.

The use of a normal 5-day working week calendar as well as the normal Finish-to-Start logic links, as per the example, provides a view of the appropriate ES, EF, LS, and LF dates. The column 'Total Float (Slack)' shows the 7-day TF against Activity 3 (as is also shown in the Project Management GPT output, however, without errors in the start and finish dates).

	Task Name	Duration	Start	Finish	Late Start	Late Finish	Total Slack	Prec
1	**PMI ChatGPT example**	**18 days**	**01/01/24**	**24/01/24**	**01/01/24**	**24/01/24**	**0 days**	
2	Activity 1	3 days	01/01/24	03/01/24	01/01/24	03/01/24	0 days	
3	Activity 2	3 days	04/01/24	08/01/24	04/01/24	08/01/24	0 days	2
4	Activity 3	4 days	04/01/24	09/01/24	15/01/24	18/01/24	7 days	2
5	Activity 4	8 days	09/01/24	18/01/24	09/01/24	18/01/24	0 days	3
6	Activity 5	4 days	19/01/24	24/01/24	19/01/24	24/01/24	0 days	5,4

Figure 8. The example in a barchart view using MSProject software.

Discussion

The results presented from a number of AI tools indicate a clear misunderstanding of CPA theory, including basic calculations and important outputs such as TF, ES, EF, LS, and LF.

If this happens with a relatively simple case, how could we trust AI's outputs to the CPA of more involved schedules with more activities and more intricate logic links, as in real life. In reality, we also introduce leads and lags, different calendars for different resources at different locations, etc., and the process of scheduling and producing a CPA output is much more intricate.

The results presented above do not allow for any confidence in the AI tool output(s), regardless of the tool used.

Another point to consider is the process between understanding the requirement, raising a query in any AI tool, and using the output. According to the PMI there should always be a '*Human in the Loop*'. Using AI requires a complex iterative process where prompts must be set out clearly and accurately and outputs thoroughly reviewed and tested before they are released to the audience.

Why are we and should we be using LLM to carry out intricate duties, such as CPA?

Over the last few decades, we have developed appropriate software tools to deliver CPA outputs for projects of any type. What is the use case for re-inventing the wheel and possibly using the 'wrong' AI model or using AI in inappropriate ways to reproduce these outputs? How can an LLM model understand schedule network outputs that cannot, at the moment, be calculated and assessed?

Shouldn't we be introducing the appropriate ML methodology within the existing scheduling software and therefore enable the AI tool to 'learn' from other similar schedules? Even better, what if we enable the ML tool to access previous real-life outputs, from real project schedules, conduct analysis of the relevant schedules and then 'instruct' it to provide us with possible outputs/scenarios?

Therefore, the author considers as more appropriate that AI tools / technology should be developed in a way to enable them to work from within the scheduling software tool(s). ML models (LLM in this case) that exist for generic purposes should not be used to provide outputs to specialist questions.

Similarly, AI implementations can be integrated with other project management software tools, for example, estimating, cost management, and contract management, to enable scenario building as well as improved performance and output from the relevant processes, rather than allowing them to bypass them.

Conclusion

We are currently in the 4th Industrial Revolution (4IR) era, and changes, as predicted by all bodies, are rapid and technically challenging. Contributing to the challenge is the scale at which various AI LLM tools have been rolled out within in the last two years and our ability to grapple with this. Despite AI having been a part of our lives for decades, the seeming 'uncontrollability' of it has caused mixed reactions in terms of the effects in our professional as well as personal life.

The issue that we are faced with is how we can best, and most of all ethically, utilise the power of the tool(s) we have developed in order to deliver an improved output. By cutting corners and bypassing processes we have developed we will not achieve a coherent output. We need to muster our collective expertise and build tools that integrate AI using existing capabilities.

The power of AI can be harnessed by applying it to support the relevant tools and not by superimposing it. Asking LLM models to deliver a Critical Path Analysis is neither appropriate nor correct. However, integrating ML tools, and perhaps a type of LLM model(s), within the scheduling software tool, or any other relevant project management software tool, and then asking it to deliver possible scenarios, is the correct approach to surf the 4th IR wave. We will still have to have the '*Human in the Loop*' to interrogate the output and advice, and therefore our next steps must be to build, educate and inform skilled personnel on how to achieve this.

References

PMI (2023). Generative AI Overview for Project Managers – Resources

Khodakarami, V., Fenton, N. E., & Neil, M. (2007). Project scheduling: improved approach to incorporate uncertainty using Bayesian networks. Project Management Journal, 38(2), 39–49.

About the Author

Dr. Dimitris N. Antoniadis

London, UK

Dr Dimitris N. Antoniadis PhD MSc BEng(1st) CEng FAPM FCMI MIMechE, based in UK, has 35+ years' experience in Programme and Project Management positions having covered project phases from concept to handover and operation / maintenance.

He is currently Director in the Programme, Project Management and PMO with DANTON PROGM, technical advisor to Novacept and having set up the BSC in Project Control he is currently the Course Leader for the partnership between London Metropolitan College and the University of West London.

He held Senior Management posts in major utilities, infrastructure and construction organisations delivering programmes of works ranging from £250M to £3.2Bn. As Head of Programme Management Office (PMO) he has set up and run the departments within challenging partnering environments, setting up all the processes from governance to reporting. He has also led / co-led major business transformation programmes for Client organisations in UK and abroad integrating project management software tools with ERP systems.

He is the author of the book '*Demystifying Project Control*'; contributed chapters in books on complexity, leadership and other project management topics and has written a number of journal and conference papers. He has been a guest speaker at UK Universities as well as International conferences on various project management topics.

He was awarded the PhD, from Loughborough University, UK, on the subject of '*Managing Complexity in Project Teams*', where he developed a framework for managing the effects of complexity on projects.

Parts of his work can be seen in www.danton-progm.co.uk

Dr. Antoniadis can be contacted at dnanton00@gmail.com