

Exploring LLM Support for Generating IEC 61131-3 Graphic Language Programs

Yimin Zhang

*CISTER / Faculty of Engineering
University of Porto
Porto, Portugal*

0009-0005-0746-315X

Mario de Sousa

*Faculty of Engineering
University of Porto
Porto, Portugal*

0000-0001-7200-1705

Abstract—The capabilities demonstrated by Large Language Models (LLMs) inspire researchers to integrate them into industrial production and automation. In the field of Programmable Logic Controller (PLC) programming, previous researchers have focused on using LLMs to generate Structured Text (ST) language, and created automatic programming workflows based on it. The IEC 61131 graphic programming languages, which still has the most users [17], have however been overlooked.

In this paper we explore using LLMs to generate graphic languages in ASCII art to provide assistance to engineers. Our series of experiments indicate that, contrary to what researchers usually think, it is possible to generate a correct Sequential Function Chart (SFC) for simple requirements when LLM is provided with several examples. On the other hand, generating a Ladder Diagram (LD) automatically remains a challenge even for very simple use cases. The automatic conversion between LD and SFC without extra information also fails when using prompt engineering alone.

Index Terms—Large Language Model, Prompt Engineering, Programmable Logic Controller, IEC 61131, Ladder Diagram, Sequential Function Chart

I. INTRODUCTION

The rise of ChatGPT and other Large Language Models (LLMs) is changing people's work and lifestyle. The same is true in industry. More and more people began to discuss what role LLMs can play in industrial domain. The most optimistic estimates even claim that a breakthrough will be achieved within a few years.

A. LLMs and Industry

LLMs can liberate engineers from heavy paperwork, allowing them to focus on their work. A fine-tuned model within an enterprise may replace some after-sales engineers, undertaking preliminary troubleshooting tasks. In the field of Artificial Intelligence (AI), these systems are often mentioned as “agents”. LLM is on par with humans in general purpose programming languages [21], which has made some engineers worried about unemployment. Several vendors have already introduced support for AI in the context of industrial automation (Siemens [20], CODESYS [19], Beckhoff [18], etc.).

Unlike the daily use of LLMs, industrial applications face hard requirements and place more emphasis on precision, clear instructions and reliability. Every procedure and production process needs to be carefully designed to avoid damage. While

we believe that LLMs can provide domain knowledge, we cannot fully rely on them. In other words, a LLM based tool is best used as a helpful assistant. Their outputs need to be rigorously validated.

B. LLMs and PLC programming

Three of the five programming languages defined in IEC 61131-3 are graphic languages - Ladder Diagram (LD), Sequential Function Chart (SFC) and Function Block Diagram (FBD). LD dominates among the 5 languages, accounting for more than 80% of the global use [17]. Its close resemblance to electrical circuits makes LD popular. Technicians who are familiar with electrical circuits but lack a knowledge of general-purpose computer languages, such as C, Java, or Python, use these graphic languages to develop PLC software.

However, LLM is typically considered a language model, which means that it's better to process textual information. The graphical nature of LD, SFC, and FBD limit the application of LLM in PLC programming. Recent researchers ([8], [10], [9]) have therefore focused on Structured Text (ST), while overlooking the most commonly used LD, SFC, and FBD (instruction List (IL) is rarely used in practice.)

However, the standard does specify these graphic languages in terms of ASCII art [16]. In actual industrial use IEC 61131-3 editors offer GUIs (Graphical User Interfaces) that rely on graphically similar elements. Nevertheless, the most important is the logic behind the graphic languages rather than the implementation form. In this paper we explore the assistance that LLMs can provide in PLC graphical programming starting from the ASCII art. In so doing we found that LLM can generate semantically and syntactically correct SFCs for simple control logic.

The rest of the paper is structured as follows. In Section II, we review the related work. We describe the methodology in Section III. Section IV showcases the experiments we did, and Section V concludes the paper.

II. STATE OF THE ART

Academic or industrial teams have released many LLMs for code generation, e.g., GitHub Copilot [15], CodeT [14], Gemini [13], Code Llama [12]. Some of them have achieved good performance in relevant tests. However, these LLMs usually

target towards general-purpose programming languages, such as Python and Java, and rarely involve PLC programming.

The team from ABB [10] systematically examines GPT-4's capability to generate ST code. They build up a preliminary benchmark containing 10 possible scenarios for industrial applications that could be used for further LLM comparison. Each category contains ten different problems. To evaluate the quality of GPT-4's answers, they designed a scoring system, measuring prompt difficulties, syntax correctness, etc. They summarize and collect the results in a workbook, providing a preliminary benchmark.

Through their experiments, the team reached some interesting conclusions. We will cite several key conclusions that are important to this study as follows:

- ChatGPT can generate syntactically correct ST;
- ChatGPT can provide domain knowledge;
- Prompt fidelity largely determines answer fidelity;

However, this research has some limitations:

- Lack of basic strategies to get better results such as providing system / assistance messages.
- Most of the prompts are zero-shot learning, without attempting few-shot learning.

In [8], Abderrahmane et al. propose a workflow, using LLM, Reinforcement Learning (RL), Natural Language Processing (NLP), etc., to automate PLC programming in Structured Text. However, they did not provide sufficient experiments and data, leaving such a pipeline still in the conceptual stage.

On the contrary, in [9], through a series of experiments, Fakih et al. demonstrated an LLM-augmented workflow that automates PLC programming also in Structured Text. The workflow includes syntax checking, formal verification, and human supervision to improve the accuracy and usability of automatically generated code. They test the pipeline on GPT-3.5, GPT-4, Code Llama-7B, etc., finally achieving an improvement of generation success rate from 47% to 72%. However, they test on a standard library, raising the concerns of the model memorizing the standard library corpus.

III. METHODOLOGY

The purpose of our experiments is to explore LLMs' capabilities in generating PLC graphic programming languages (LD, SFC and FBD). Due to the nature of our research interests, for the moment we have focused our experiments on situations better modeled as discrete event systems, for which FBD is not the most appropriate language.

A. Prompts / Questions

The experiments conducted by [10] revealed that ChatGPT can generate syntactically correct IEC 61131-3 Structured Text code. A very basic test to generate SFC was not successful. For the sake of comparison we continue to use the benchmark¹ they provided in the first experiments. We limited our research to the Categories 3 to 5 that are related to discrete event systems i.e. PLC Programming Exercises, Process Control and

Sequential Control. The prompts used from Experiment 3 to Experiment 6 will be discussed later in Section IV-C.

B. LLM Model

Considering its popularity and ease of comparison, we choose OpenAI API, "gpt-4-turbo-preview", for our experimentation. The version is "gpt-4-0125-preview" for which OpenAI claims that "this model completes tasks like code generation more thoroughly than the previous preview model and is intended to reduce cases of 'laziness' where the model does not complete a task [6]."

C. Few-shot Learning

Previous preliminary prompt engineering, which could be considered as zero-shot learning, has already demonstrated some surprising conclusions in PLC programming. However, LLMs are few-shot learners [5]. It is generally believed that the performance of few-shot learning is better than zero-shot learning, which prompts our curiosity: can LLMs handle more complex PLC programming when given examples?

D. Evaluation

At present, we are not aware of any compiler that can address programs in ASCII art format, even though the IEC 61131-3 standard itself formally defines ASCII art representations for all graphical languages, including LD. It is therefore difficult to automate the evaluation of the correctness of a LLM's outputs. One approach is to manually input LLM's answers into an IDE, but this would entail a huge workload and go against the initial purpose. The method we adopt is manual judgment and scoring of the answers, which is also made possible by the limited number of discrete events in the problems being analysed.

For transparency all outputs from LLM interface are published on GitHub².

TABLE I
SCORING INSTRUCTIONS

Score	Meaning
-1	Don't provide a solution in required language.
0	Semantically or syntactically Wrong.
0.5	Wrong, but engineers can understand.
1	Semantically and syntactically correct.

Regarding the scoring, the Table I outlines the rules for scoring. It should be noted that in order to avoid giving abrupt judgments of incorrect or correct answers, we divide this into 4 levels. Half points represent that, although not syntactically and semantically correct, engineers with basic knowledge of LD or SFC can understand the intention of the graphics. As mentioned in the introduction, this can also serve as programming assistance.

IV. EXPERIMENTS

A summary of all experiments is shown in Table II describing strategies applied in each experiment.

¹<https://github.com/hkoziolek/control-logic-generation-prompts>

²https://github.com/yimin-up/LLM_IEC-61131_Graphic

TABLE II
EXPERIMENTS SUMMARY

Index	System Message	Subtask	Few-shot
Experiment 1	✓	✓	✓
Experiment 2	✓	✓	✓
Experiment 3	✓	x	✓
Experiment 4	✓	x	✓
Experiment 5	✓	x	✓
Experiment 6	✓	x	✓
Experiment 7	✓	x	✓

A. Experiment 1: PLC programs in Ladder Diagram

In this experiment, we require GPT-4 model to provide the Ladder Diagram solutions for 30 questions (10 questions from each of 3 categories). We use the scoring mechanism mentioned earlier and summarize the results in Table III. Due to space constraints only the 10 questions of Category 3 are displayed here for ease of comparison.

We will compare the following three strategies:

- System messages: *You are an AI expert in PLC (Programmable Logic Controller) programming. You are familiar with the international standard IEC 61131 for PLCs. You will be provided with some design requirements. Your task is to write IEC 61131-3 Ladder Diagram programs in ASCII art according to the requirements.*
- System messages divided into steps: *Follow these steps to fulfil the requirements:*
 - *Step 1: Define and list the inputs and outputs of the Ladder Diagram, and the standard symbols they use.*
 - *Step 2: Give the Ladder Diagram program.*
 - *Step 3: Give the logical expressions for each rung.*
 - *Step 4: Briefly describe the state machine of the Ladder Diagram.*
 - *Step 5: Check the LD you generate according to the IEC 61131-3 standard.*
- Few-shot learning: provide 3 examples in prompts.

The 3 examples are:

Example 1

User messages: *Use start / stop button to control a motor. When the start button is pressed, the motor starts. When the stop button is pressed, the motor stops.*

Assistant messages:

Below is a simple representation of the ladder diagram in ASCII art.

```
```
(Start of Ladder Diagram)
| |
| StartButton StopButton Motor |
+---+ |-----+-----| / |-----()---+
| | |
| Motor | |
+---+ |-----+ |
(End of Ladder Diagram)
```
```

Example 2

User messages: *Design a counter with reset control input. If reset is True, counter value is reset to the value defined by global configuration constant ResetCounterValue. If reset is False, counter is incremented every PLC cycle.*

Assistant messages:

The Ladder Diagram program in ASCII art:

``

(Start of Ladder Diagram)

```
+-----+           +-----+
|           | SEL   |           |
| Reset   |           |           |
+---+ |-----+-----| G     OUT |--cnt |
|           |           |           |
+-----+           +-----+
|           | ADD   |           | | |
| 1--| IN1 OUT |----| IN0   |           |
| cnt--| IN2   |           |           |
+-----+ +--+| IN1   |           |
|           | +-----+           |
| ResetValue--+
(End of Ladder Diagram)
```
```

#### Example 3

User messages: *Create a program for 8-bit running LED lights.*

Assistant messages:

Ladder Diagram Program:

``

(Start of Ladder Diagram)

```
+-----+ status |
| start |-----(S)---+
| |
| stop status |
+---+ |-----+-----+-----(R)---+
| |
| +-----+
| status | TON | |
+---+ |-----| EN ENO |--+
| | |
| NOT delay.Q --| IN Q |--+
| T#1S --| PT ET |--+
| +-----+
| +-----+
| delay.Q | ROL | |
+---+ |-----| EN ENO |--+
LED8 --	IN ERG	-- LED8
1 --	N	
+-----+		
(End of Ladder Diagram)
```
```

1) *Results:* The results are summarized in Table III.

TABLE III
SCORES FOR GPT4 GENERATED LD

Number	CaseShortName	System Message	Step	Few-shot
3_1	ConveyorControl	0	0	0.5
3_2	HeatingControl	0	0	0
3_3	TrafficControl	0	0	0
3_4	PneumaticControl	0	0	0
3_5	ElevatorControl	0	0	0
3_6	CarWash	0	0	0.5
3_7	CarPark	0	0	0
3_8	PickPlace	0	0	0
3_9	BottleRemoval	0	0	0.5
3_10	CoffeeMaker	0	0	0

2) *Analysis:* In the case of existing cases, no correct LD has been generated. In few-shot learning, however, LD that is understandable by humans has been generated. To be more precise, it can be understood by engineers who have basic PLC programming knowledge. Upon deeper analysis, we found that these cases are relatively simple cases. Therefore, we have designed several more cases in subsequent sections (Experiment 3 and 4), simplified the logic, and provided more design details.

The results also demonstrate that GPT-4 struggles with understanding IEC 61131-3, because there are many errors in the symbols it provides. Many symbols do not exist in the standard, which means that GPT-4 has “hallucinated” these symbols.

One important reason for failure is that spaces are not counted as a token. We tested this on the tokenizer tool provided by OpenAI³. Adding or reducing spaces does not affect the number of tokens; it only affects the character count. However, spaces do affect ASCII art diagram, which impacts people’s understanding of the LD. Another reason may be that there is no specific fine-tuning specifically for ASCII art.

3) Lessons Learned:

- The performance difference between LD and ST for the same problem is like night and day. To simplify the requirements, we will try to propose more detailed specifications. Specifically, we’ll abandon complexity, including elements such as timers, counters, etc., that may be difficult to understand for machine.
- Few-shot learning helps improve the results in some cases. For simple logic tests in Experiment 3 and 4, we will provide simple examples too.

B. Experiment 2: PLC programs in Sequential Function Chart

In this set of experiments, we repeat the experiments in Experiment 1. The difference is that the programming language will be changed to SFC. In task decomposition, some requirements will be modified according to SFC language. Due to space limitations, they will not be displayed here. In few-shot learning we also provide 3 examples, which are not demonstrated here for the same reason.

1) *Results:* The results are summarized in Table IV.

³<https://platform.openai.com/tokenizer>

TABLE IV
SCORES FOR GPT4 GENERATED SFC

Case	CaseShortName	System Message	Step	Few-shot
3_1	ConveyorControl	0	0	-1
3_2	HeatingControl	0	0	0.5
3_3	TrafficControl	0	0	0
3_4	PneumaticControl	0	0	0
3_5	ElevatorControl	0	0	-1
3_6	CarWash	0	0	0.5
3_7	CarPark	0	0	0
3_8	PickPlace	0	0	0
3_9	BottleRemoval	0	0	0.5
3_10	CoffeeMaker	0	0	0

2) *Analysis:* The following analysis is based on few-shot learning. Case 3_1 and 3_5 does not provide SFCs at all, they only describes SFCs in text, which does not comply with IEC 61131-3. Case 3_5 and 3_8 try to identify synchronous paths, but failed. Case 3_6 and 3_9 produced something resemble flowcharts rather than SFCs.

As for the reason, we believe it is similar to the previous experiment, namely that spaces are not counted as tokens.

3) Lessons Learned:

- For complex logic, GPT-4 cannot generate correct or useful SFCs.
- Few-shot learning helps improve the results.
- We need to find a way to treat spaces as valid tokens.

C. Experiment 3: LD - Detailed Cases

In the previous experiments, whether using GPT-4 to generate LD or SFC, it can be considered as unsuccessful. It seems that GPT is not competent enough for complex logic. Therefore, we simplified the logical requirements of the design, removing potentially challenging logic such as timing and counting, and only examined the situation of discrete event logic. There is also no parallel operation involved. Furthermore, we provided detailed names for each sensor and action logic.

Based on the conclusions from our previous experiments, decomposing tasks did not significantly increase accuracy, while few-shot learning significantly improved the outputs. Therefore, under these experimental conditions, we abandon task steps and adopt the combination of system messages with few-shot learning.

For this and subsequent experiments (i.e. Experiment 3 to 6), we designed 3 test cases that focus on discrete event systems with a limited number of states. While in Case 1 the sequence is explained sequentially in the text (making it easier to interpret and convert to a discrete event based program), Case 3 has the sequence implicitly defined by the objectives that need to be achieved (making it more difficult to convert to a program). In terms of size complexity, case 1 involves two-bit logic while Case 2 involves only one-bit logic. Case 3 can be regarded as involving four-bit logic with 2 hidden states.

The prompts given to the LLM tool are provided below. For sake of comparison with LLM-generated code, we also provide

examples of what we would consider a correct solution for Case 1: using LD in Figure 1, and SFC in Figure 2.

Case 1: Press Control You want to control a hydraulic press. Consider the following specification:

- The system performs an operating cycle every time the start button (BI) is pressed.
- The press starts the cycle in the upper position (detected by sensor S going high), descending until it reaches the lower position (detected by sensor I going high).
- After the press reaches the lower position, the cycle continues by moving up until reaching the upper position, ending the cycle.
- The press is driven by a motor M with movement control in both directions: downward (M+) and upward (M-).

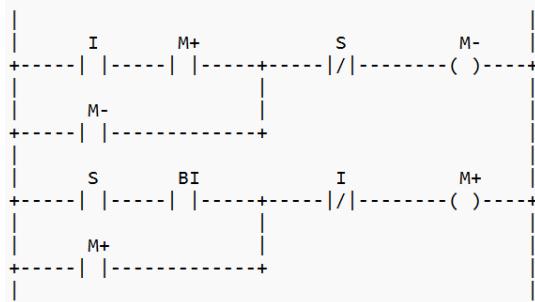


Fig. 1. LD solution for Case 1

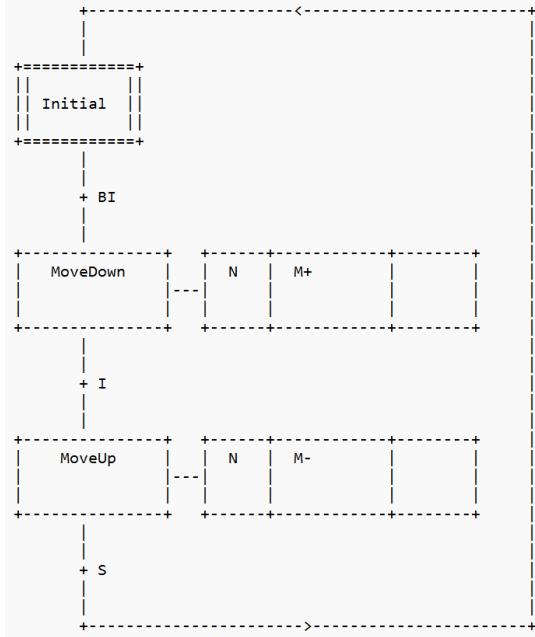


Fig. 2. SFC solution for Case 1

Case 2: Motor Start / Stop Use start / stop button to control a motor. When the start button is pressed, the motor starts. When the stop button is pressed, the motor stops.

Case 3: Two Pumps There are two water pumps (P1, P2) that pump out the water in a reservoir. To extend the lifespan

of the water pumps, the two pumps operate alternately. When the water level is above the high water line (HL), the pump operates. When the water level is below the low water line (LL), the pump stops working.

1) Results & Analysis: Due to the limited number of cases, we are not using a scoring mechanism. Actually, no correct LD solutions were generated - the severely mangled generated output attempting to mimic LD would take up considerable space, so we will not list the results here. For the complete results, please refer to the GitHub repository⁴.

The experimental results indicate that GPT-4 still cannot generate correct LD even for simple scenarios. This suggests that LD poses a significant challenge for LLMs.

D. Experiment 4: SFC - Detailed Cases

This experiment used the same cases and procedures as the previous Experiment 4, but now asking for results in SFC language.

1) Results: For Case 1 and 2, GPT-4 yielded completely semantically and syntactically correct results. In Case 3 the hidden states were not generated, resulting in an incorrect outcome. All results will be listed here (Figure 3 - Figure 5). The correct SFC for Case 3 is shown in Figure 6 for comparison.

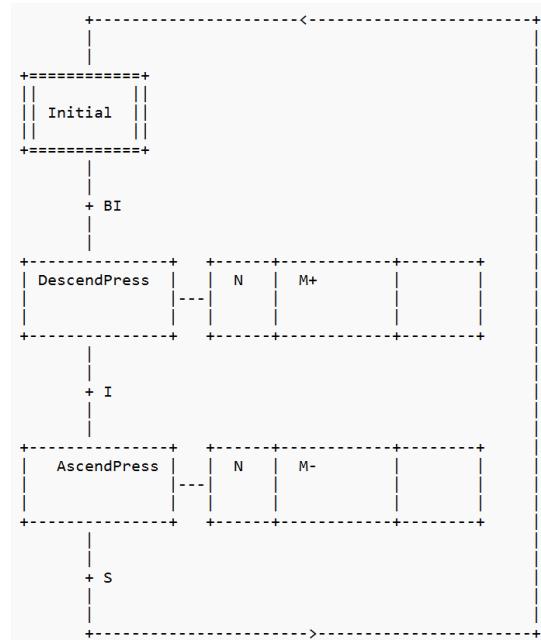


Fig. 3. SFC solution generated by GPT-4 for Case 1

2) Analysis: These outcomes are unexpected. They indicate that SFC is more easily understood and learned by GPT-4. However, if we repeat the same prompts several times, we get different outputs which may be incorrect. Especially for Case 3, no correct answer was generated. But for Case 1 and 2, containing 1-bit or 2-bit state machine, the output is correct.

⁴https://github.com/yimin-up/LLM_IEC-61131_Graphic

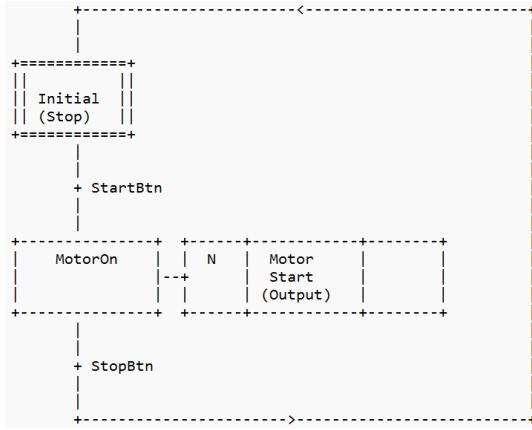


Fig. 4. SFC solution generated by GPT-4 for Case 2

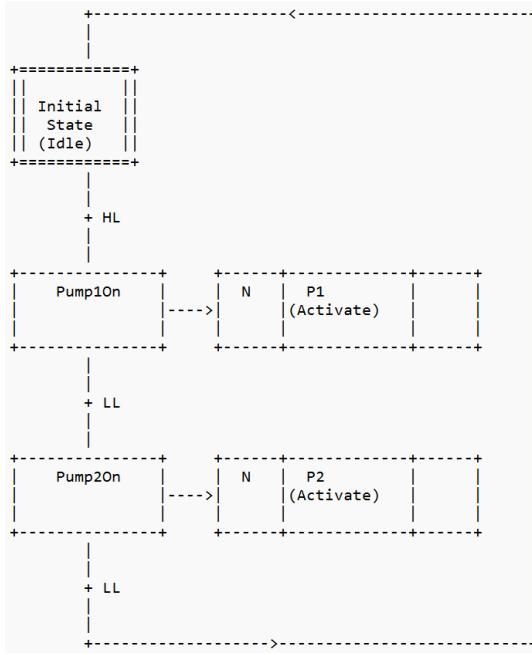


Fig. 5. SFC solution generated by GPT-4 for Case 3

E. Experiment 5: LD-SFC Conversion

The conversion problem dates back to at least the 1990s when [4] described it as a design recovery problem in 1992. Since then, a few algorithms were proposed to convert LD into SFC, including graphical analysis [4], temporal logic methods [2], state-space based algorithms [1], etc. In practice, these algorithms have not been widely adopted because they cannot solve issues such as a lack of domain knowledge and state space explosion.

In this experiment, we provide three conversion examples in the prompts, from LD to SFC, without any textual description of the problem so as to guarantee that the conversion doesn't come from the textual information. The three examples are simple sequential sequences that loop back to the initial state, with 2, 3 and 4 states each. All examples provided use the

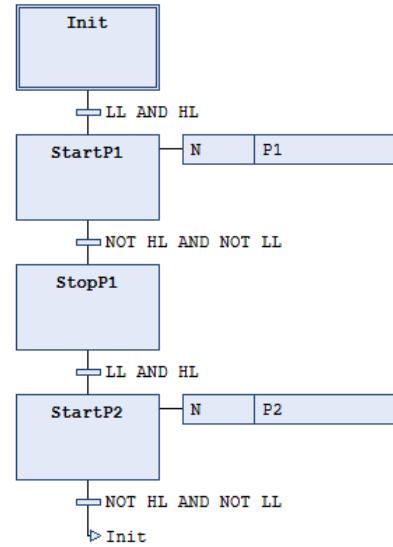


Fig. 6. SFC solution in CODESYS for Case 3

same design pattern for implementing state machines using LD diagram, as exemplified in Figure 1.

We then provide the LD solution to each of the three test cases, and ask for the equivalent solution in SFC. It should be noted that the test cases are in essence identical to the examples given, with test case 1 having 3 sequential states, test case 2 having 2 states, and test case 3 having 4 sequential states. Once again we use the same design pattern for implementing state machines in LD diagram. These three test cases are however different to the examples given when taking into account the names of the variables (sensors and actuators). One of the examples also has some extra outputs (warning lights) that distinguishes it from the canonical 3-state sequential problem.

1) Results & Analysis: No correct conversion was achieved. The results will not be listed here. For more information, please refer to the GitHub repository. The results produced for the three examples resemble more of a flowchart, but even so incorrect in terms of identifying the number of states and their sequence; perhaps GPT-4 tends to generate flowcharts more often.

The difficulty in conversion may lie in three aspects:

- Poor understanding of LD as Experiment 1 and 3 showed.
- The absence of auxiliary contextual information. As a language model, LLMs understand text better than ASCII art or graphical information.
- Lack corresponding corpora to training the model.

All of these aspects are worth further study.

F. Experiment 6: SFC-LD Conversion

Contrary to the previous Experiment 5, there are rule-based solutions for converting a SFC into a LD. Theoretically, it's less challenging. We once again provide the same three conversion examples in the prompts without any descriptions,

and ask to convert the SFC solution to each of the test cases into equivalent LD programs.

1) Results & Analysis: Unfortunately, no correct conversion was achieved. Surprisingly, even for test case 2, which is a simple case with only one logical operation, resulted in a completely wrong output. Although the graph generated for test case 3 is incorrect, it leads to the deduction of four states with 2 hidden state. It shows GPT-4 can understand the SFC in this case compared to Experiment 5. For this case, the correct LD could be obtained with slight manual modifications. Due to space constraints, we refer the reader to the GitHub repository for the complete results. As for possible reasons, we believe it's similar to the previous conclusion in Experiment 5.

G. Experiment 7: GPT-4 vs Claude 3 vs Gemini

This experiment will compare results obtained when using the currently popular LLMs: GPT-4, Claude 3 and Gemini-1.5. We adopt the few-shot learning strategy and use the scoring system mentioned earlier. For ease of comparison, we tested on Category 3 (PLC Programming Exercises) using APIs.

1) Results: Table V shows the results from different LLMs.

TABLE V
COMPARISON BETWEEN GPT-4, CLAUDE 3 AND GEMINI-1.5

Case	GPT-4		Claude 3		Gemini-1.5	
	LD	SFC	LD	SFC	LD	SFC
3_1	0.5	-1	0	-1	0	0
3_2	0	0.5	-1	-1	0.5	0.5
3_3	0	0	-1	-1	0	0
3_4	0	0	0	-1	0	0
3_5	0	-1	-1	0	0	0
3_6	0.5	0.5	0.5	0.5	0	0.5
3_7	0	0	-1	0.5	0	0
3_8	0	0	-1	0	0	0
3_9	0.5	0.5	0	0	0	0.5
3_10	0	0	-1	0	0	0.5

2) Analysis: There is no clear winner or loser when comparing the results obtained, and all LLMs struggled to provide meaningful results. No LLM dominates the others, with each LLM able to provide a better results than the rest in at least one of the test cases. However it could be argued that Claude 3 is the weakest in this task because it more often doesn't provide a solution at all, or the solution provided is not in the required language.

V. CONCLUSION AND FUTURE WORK

We conducted a series of experiments to investigate the ability of LLMs in generating LD and SFC for PLC programming. The results indicate that LLM is more capable of understanding SFC. However, for LD, LLM finds it relatively difficult to comprehend. LLM can provide correct solutions for relatively simple state machines (1-bit, 2-bit) according to textual descriptions. However, for complex logic, it is currently incapable. Moreover, regarding the conversion between LD and SFC, LLMs are not yet competent.

Future work includes Retrieval Augmented Generation (RAG) for LLMs, and fine-tune the models, trying to improving the accuracy of SFC, the correctness of LD, and the correctness in complex tasks. Considering how to address spaces in ASCII art will be the next key focus. Currently, all these experiments are black-box testings. Trying to find theoretical explanation for these results is another interesting direction.

REFERENCES

- [1] R. Vimal Nandhan and N. Ramesh Babu, "Understanding of Logic in Ladder Program with Its Transformation into Sequential Graph Using State Space-Based Approach," *International Journal of Mechatronics and Manufacturing Systems*, vol. 6, no. 2, pp. 159–182, 2013, pMID: 53827. [Online]. Available: <https://www.inderscienceonline.com/doi/abs/10.1504/IJMMS.2013.053827>
- [2] T. Zanma, T. Suzuki, A. Inaba, and S. Okuma, "Transformation Algorithm from Ladder Diagram to SFC Using Temporal Logic," *IEEE Transactions on Industry Applications*, vol. 117, no. 12, pp. 1471–1479, 1997.
- [3] A. Falcione and B. Krogh, "Design recovery for relay ladder logic," *IEEE Control Systems Magazine*, vol. 13, no. 2, pp. 90–98, 1993.
- [4] ———, "Design recovery for relay ladder logic," in *[Proceedings 1992] The First IEEE Conference on Control Applications*, 1992, pp. 648–653 vol.2.
- [5] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language Models are Few-Shot Learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [6] OpenAI, "New Embedding Models and API Updates," January 2024, available at <https://openai.com/blog/new-embedding-models-and-api-updates>.
- [7] ———, "Prompt Engineering," available at <https://platform.openai.com/docs/guides/prompt-engineering>.
- [8] A. Boudribila, M.-A. Chadi, A. Tajer, and Z. Boulghasoul, "Large Language Models and Adversarial Reinforcement Learning to Automate PLCs Programming: A Preliminary Investigation," in *2023 9th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2023, pp. 650–655.
- [9] M. Fakih, R. Dharmaji, Y. Moghaddas, G. Q. Araya, O. Ogundare, and M. A. A. Faruque, "LLM4PLC: Harnessing Large Language Models for Verifiable Programming of PLCs in Industrial Control Systems," *arXiv preprint arXiv:2401.05443*, 2024.
- [10] H. Koziolek, S. Gruener, and V. Ashiwal, "ChatGPT for PLC/DCS Control Logic Generation," in *2023 IEEE 28th International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2023, pp. 1–8.
- [11] Anthropic, "The Claude 3 Model Family: Opus, Sonnet, Haiku," March 2024, available at <https://www.anthropic.com/clause>.
- [12] B. Rozière, J. Gehring, F. Gloeckle, S. Sootla, I. Gat, X. E. Tan, Y. Adi, J. Liu, R. Sauvestre, T. Remez, J. Rapin, A. Kozhevnikov, I. Evtimov, J. Bitton, M. Bhatt, C. C. Ferrer, A. Grattafiori, W. Xiong, A. Défossez, J. Copet, F. Azhar, H. Touvron, L. Martin, N. Usunier, T. Scialom, and G. Synnaeve, "Code Llama: Open Foundation Models for Code," 2024.
- [13] G. Team, R. Anil, S. Borgeaud, Y. Wu, J.-B. Alayrac, J. Yu, R. Soricut, J. Schalkwyk, A. M. Dai, A. Hauth *et al.*, "Gemini: A Family of Highly Capable Multimodal Models," 2023.
- [14] B. Chen, F. Zhang, A. Nguyen, D. Zan, Z. Lin, J.-G. Lou, and W. Chen, "CodeT: Code Generation with Generated Tests," 2022.
- [15] GitHub, "Copilot," September 2021, available at <https://github.com/features/copilot>.
- [16] IEC, "IEC 61131-3, 3rd Ed. Programmable Controllers – Programming Languages," February 2013, International Electrotechnical Commission.
- [17] Technavio, "Global PLC Software Market 2016-2020," April 2016, available at <https://www.technavio.com/report/global-automation-plc-software-market>.
- [18] Beckhoff, "TwinCAT Projects with AI-assisted Engineering," 2023, Available at <https://www.beckhoff.com/en-en/products/automation/twincat-projects-with-ai-supported-engineering>.
- [19] CODESYS, "Using AI to code with CODESYS," October 2023, Available at <https://www.codesys.us/codesys-tutorials.html>.

- [20] Siemens, “Autonomous AI Skills for Advanced Robotics,” Available at <https://www.siemens.com/global/en/products/automation/topic-areas/tia/future-topics/simatic-robotics-ai.html>.
- [21] J. Achiam, S. Adler, S. Agarwal, L. Ahmad, I. Akkaya, F. L. Aleman, D. Almeida, J. Altenschmidt, S. Altman, S. Anadkat *et al.*, “GPT-4 Technical Report,” *arXiv preprint arXiv:2303.08774*, 2023.