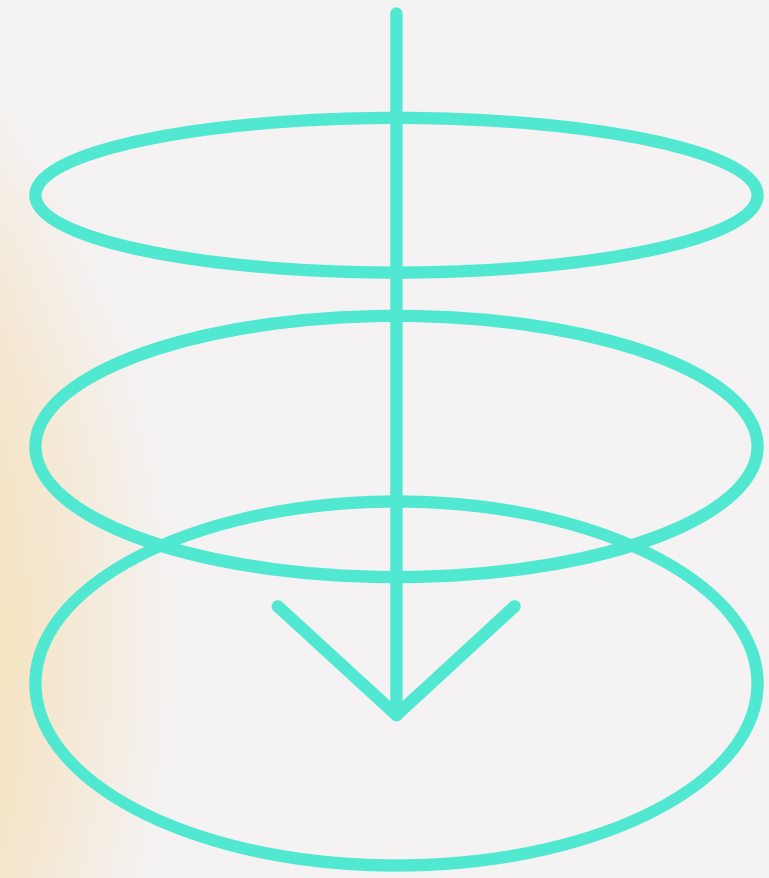


# Applied Artificial intelligence

Actuated Traffic Signal Control System using Large Language Models

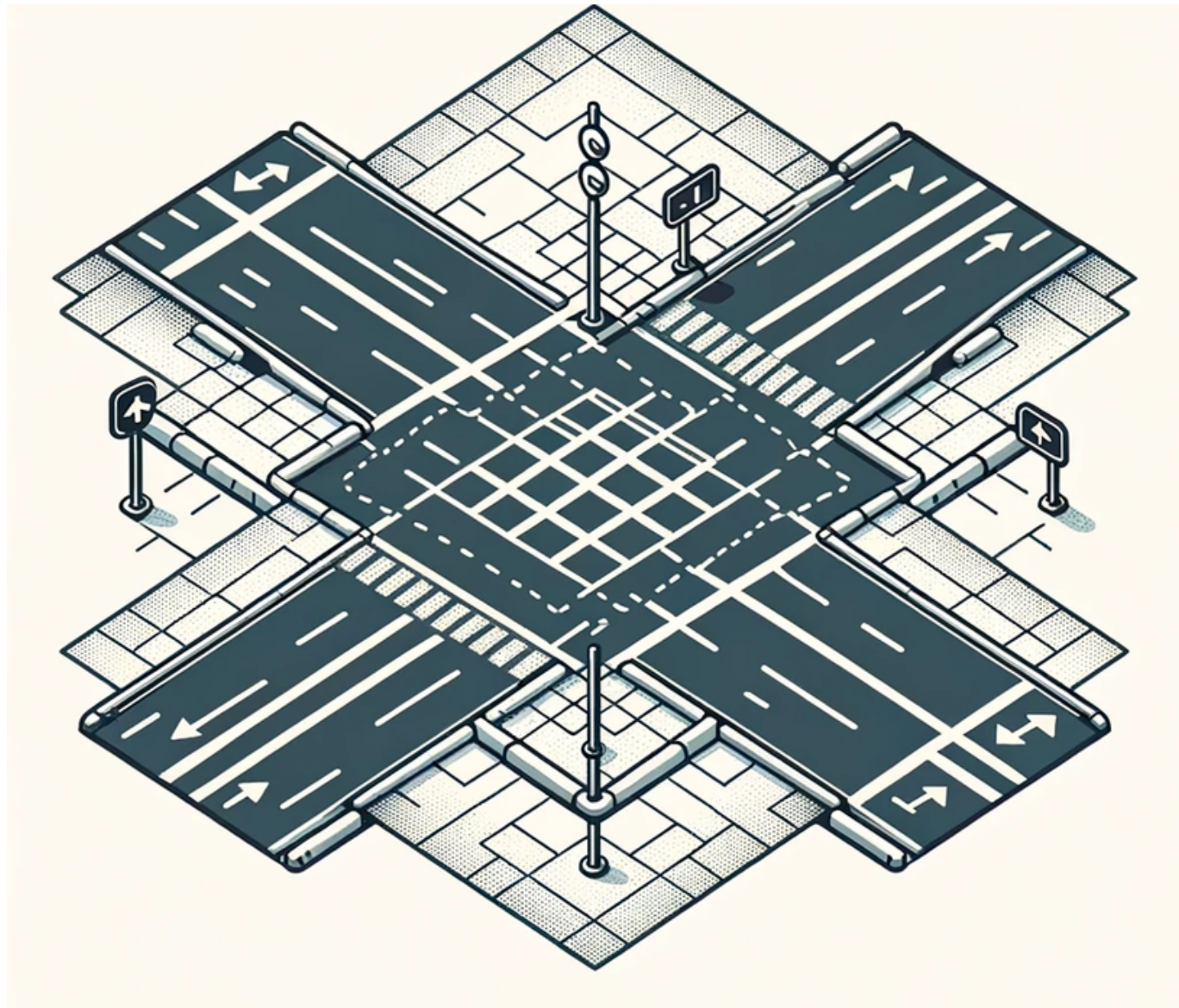


**CS514**

Applied AI

Kaushal Patil

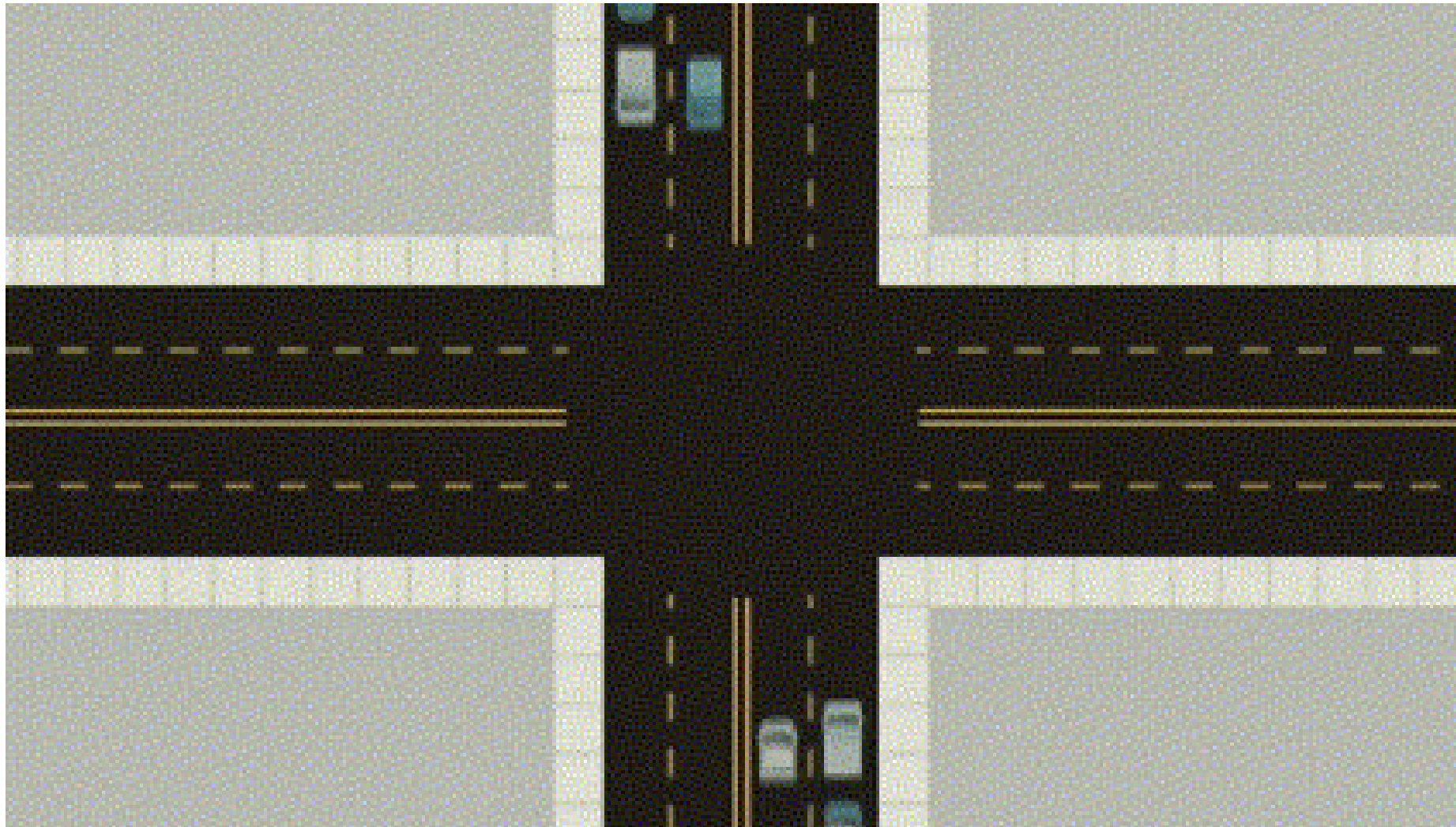
## Most Intuitive: Round Robin Control



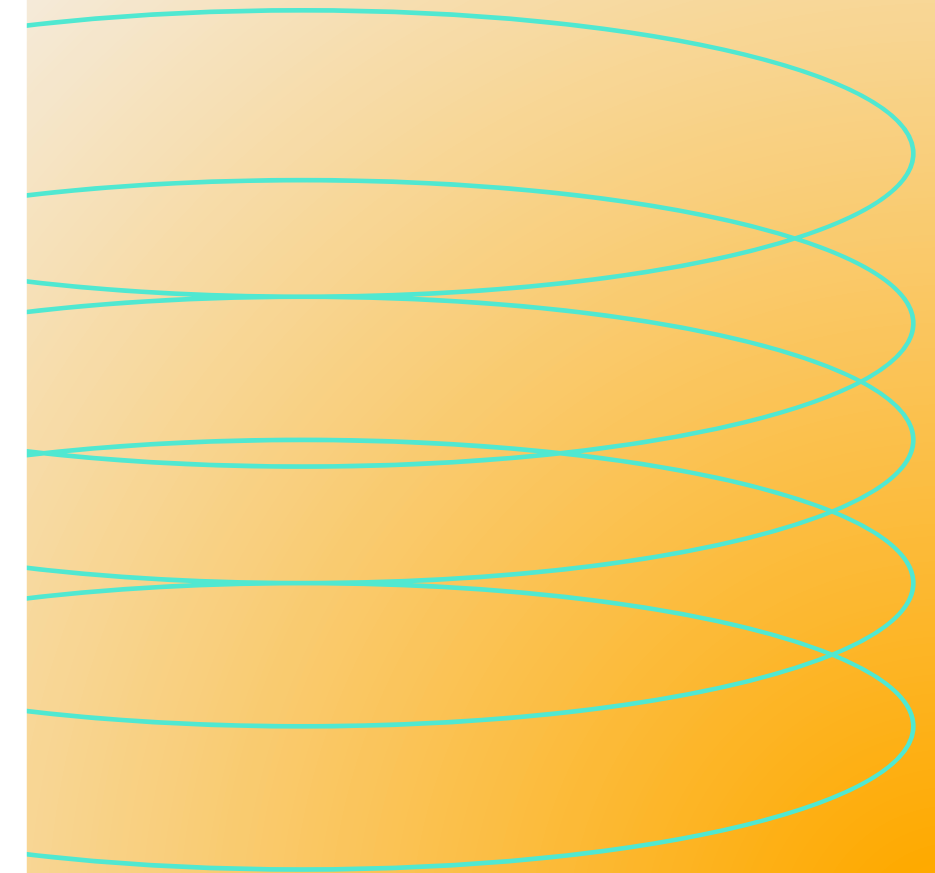
Source: Dall E

Traffic Signals

## North America



Traffic Signals





# Deciding which signal to open based on

Formulation of  
Problem

Each direction end has the following properties:

- Number of Vehicles waiting for through or right
- Number of Vehicles waiting for left
- Cycles since the last waiting vehicle arrived left
- Cycles since the last waiting vehicle arrived through or right
- Number of pedestrians waiting on parallel lane
- Cycles since the last pedestrian arrived
- Emergency vehicles for through or right
- Emergency vehicles for left

Global properties:

- Public transport vehicle timings (transit event)
- Busy streets and their timings (busy street events)

## LLMs Evaluated

GPT4

GPT3.5

Mistral Instruct 7B

Vicuna 13B

Llama 2 13B

I have evaluated a simple version of the problem using these models, where in we just try to predict which signal to open based on the number of vehicles information at each signal.

I have tried the following prompting strategies:

- Zero Shot Prompting
- Few Shot Prompting
- Chain of Thought Prompting

# Comparison Example

## (Evaluation Sheet)

LLM Name	Test Case ID	Rule Based/Bayesian Network/Influence Network Output (Output for North American System)	Adjusted Output for Task (Output if only one lane can be opened)	Predicted for North American System	Predicted for General Traffic Signals	Reasoning	Math
ChatGPT 4	#4	North South Left Lane	South Lane	North South Left Lane	South Lane	Very High	Very Accurate
ChatGPT 4	#2	East West Through Right Lane	West Lane	East West Through Right Lane	West Lane	Very High	Very Accurate
ChatGPT 3.5	#4	North South Left Lane	South Lane	North South Through Right	South Lane	High	Accurate
ChatGPT 3.5	#2	East West Through Right Lane	West Lane	North South Through Right	West Lane	High	Accurate
Mistral 7B	#4	North South Left Lane	South Lane	North South Through Right	West Lane	Low	NA
Mistral 7B	#2	East West Through Right Lane	West Lane	North South Through Right	North South Through Right	Low	NA
Vicuna 13B	#4	North South Left Lane	South Lane	North South Through Right	South Lane	Low	Accurate
Vicuna 13B	#2	East West Through Right Lane	West Lane	North South Through Right	West Lane	Low	Accurate
Llama 2 13B	#4	North South Left Lane	South Lane	South Through and Left	South Lane	Medium	Accurate
Llama 2 13B	#2	East West Through Right Lane	West Lane	East Through Right	East Lane	Low	Low

Model Name	Reasoning (out of 4)	Math (out of 4)	Prediction Capability Adj	Prediciton Capability (out of 100)	
ChatGPT4	4	4	4	100	
ChaGPT3	3	2	2.5	62.5	
Mistral 7B	1	0	0.75	18.75	
Vicuna 13B	1	2	2.75	68.75	
Llama 2 13B	1.5	1.5	2.24	56	

# Zero-Shot Learning

*Prompt:*

```
Classify the text into neutral, negative or positive.  
Text: I think the vacation is okay.  
Sentiment:
```

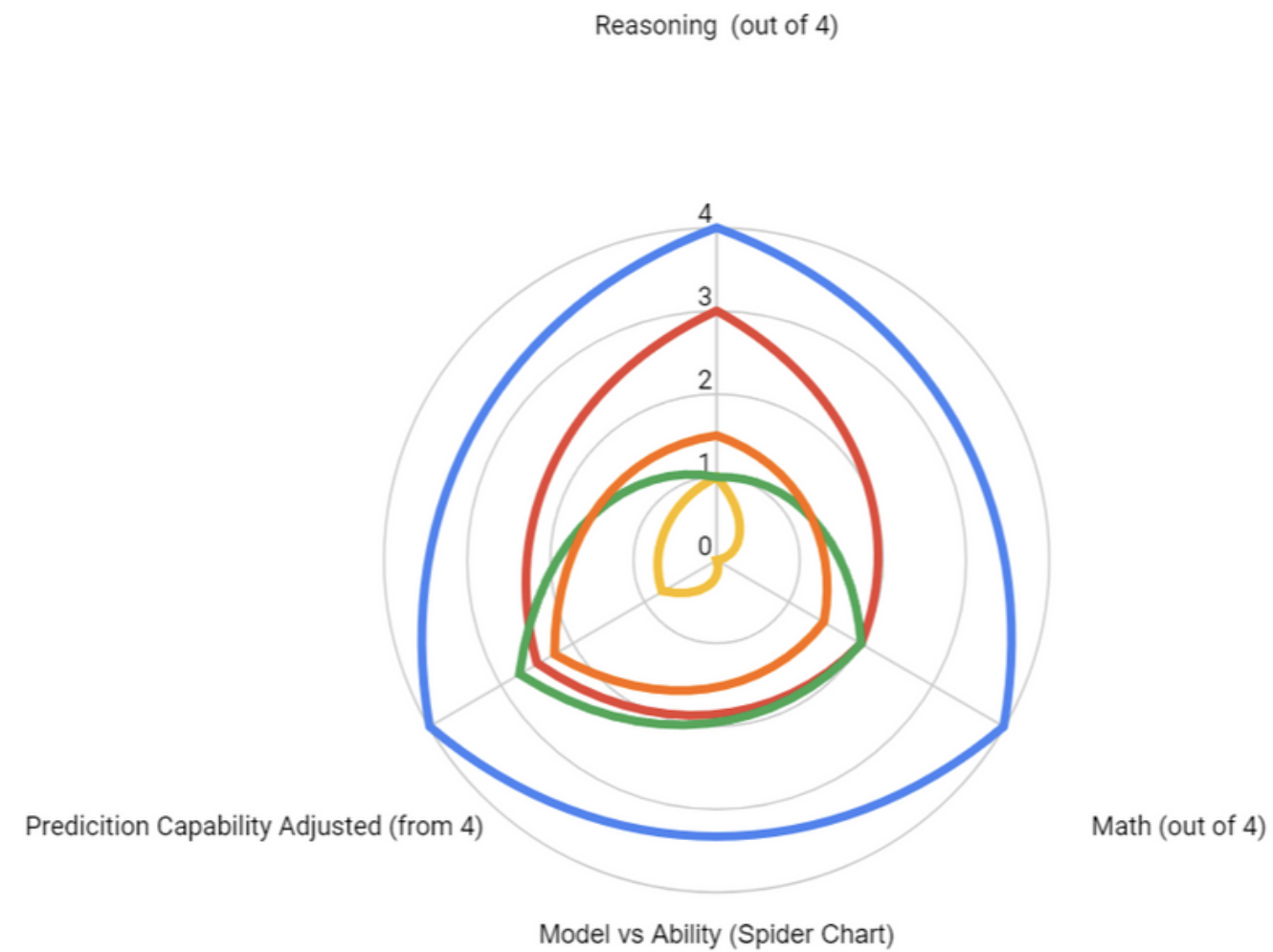
*Output:*

```
Neutral
```

# Zero-Shot Learning Prompt Lists, Evaluation Sheet

Reasoning , Math and Prediction Capability

ChatGPT4 ChaGPT3 Mistral 7B Vicuna 13B Llama 2 13B





# Few Shot Learning

## *Prompt:*

A "whatpu" is a small, furry animal native to Tanzania. An example of a sentence that uses the word whatpu is:  
We were traveling in Africa and we saw these very cute whatpus.  
To do a "farduddle" means to jump up and down really fast. An example of a sentence that uses the word farduddle is:

## *Output:*

When we won the game, we all started to farduddle in celebration.

# Few Shot Learning Prompt Lists, Evaluation Sheet

Reasoning (out of 4), Math (out of 4) and Prediction Capability Adjusted (from 4)

ChaGPT3   Mistral 7B   Vicuna 13B   Llama 2 13B  
Reasoning (out of 4)



Model vs Ability (Spider Chart ) [Few Shot Learning]

# Chain of Thought Prompting

## Standard Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The answer is 27. ❌

## Chain-of-Thought Prompting

### Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls.  $5 + 6 = 11$ . The answer is 11.

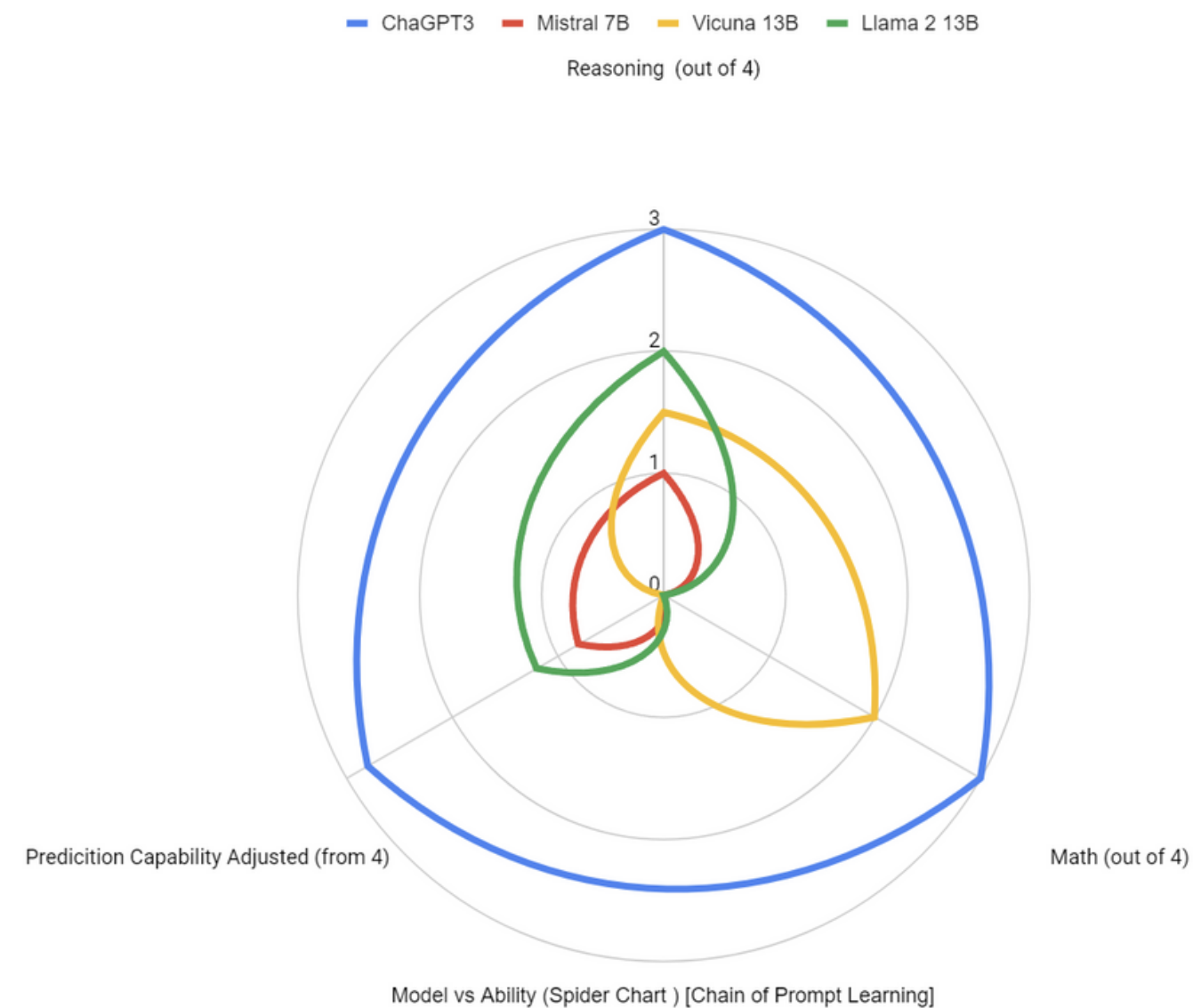
Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

### Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had  $23 - 20 = 3$ . They bought 6 more apples, so they have  $3 + 6 = 9$ . The answer is 9. ✅

# Chain of Thought Prompting Prompt Lists, Evaluation Sheet

Reasoning (out of 4), Math (out of 4) and Prediction Capability Adjusted (from 4)





Utilising GPT 4

Zero shot  
converstation with  
GPT 4

And the best model is GPT-4

# Conclusions

## Pros:

1. Versatility for managing signals in contract to rule based / bayesian network which has to be changed.
2. Able to handle busy street events, public transits, emergency vehicles, and scope to handle more unpredictable events.
3. Ability to factor in new knowlegde very easily

## Cons:

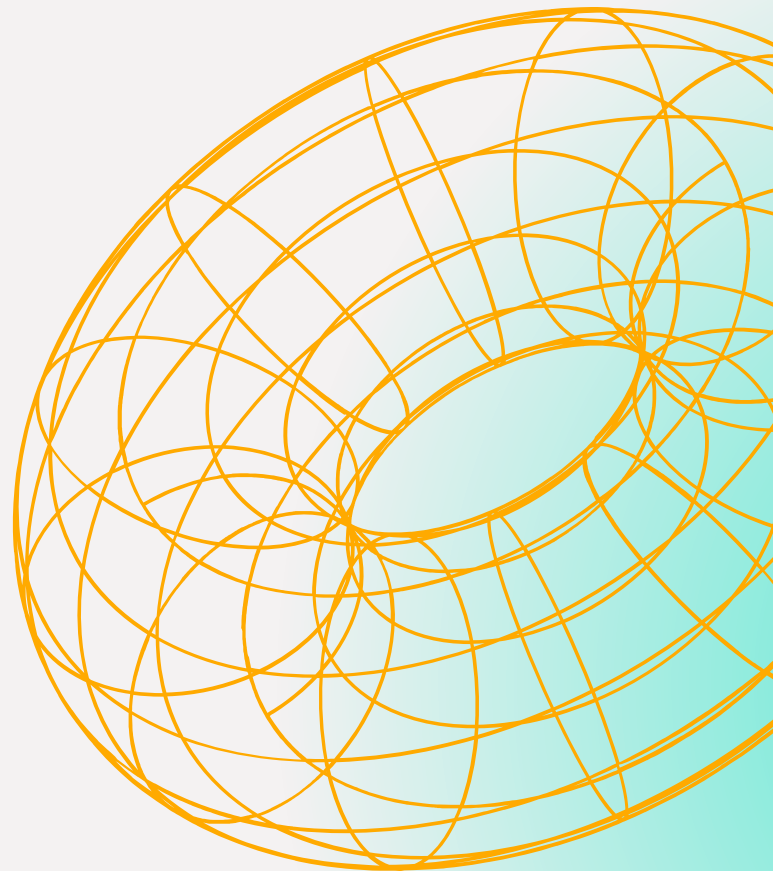
1. Temperature based predictions leads to unpredictable output formats.
2. GPT4 is a very big model and using it for traffic management of a single intersection may be impractical ( for today's time)

Why GPT 4 over other models

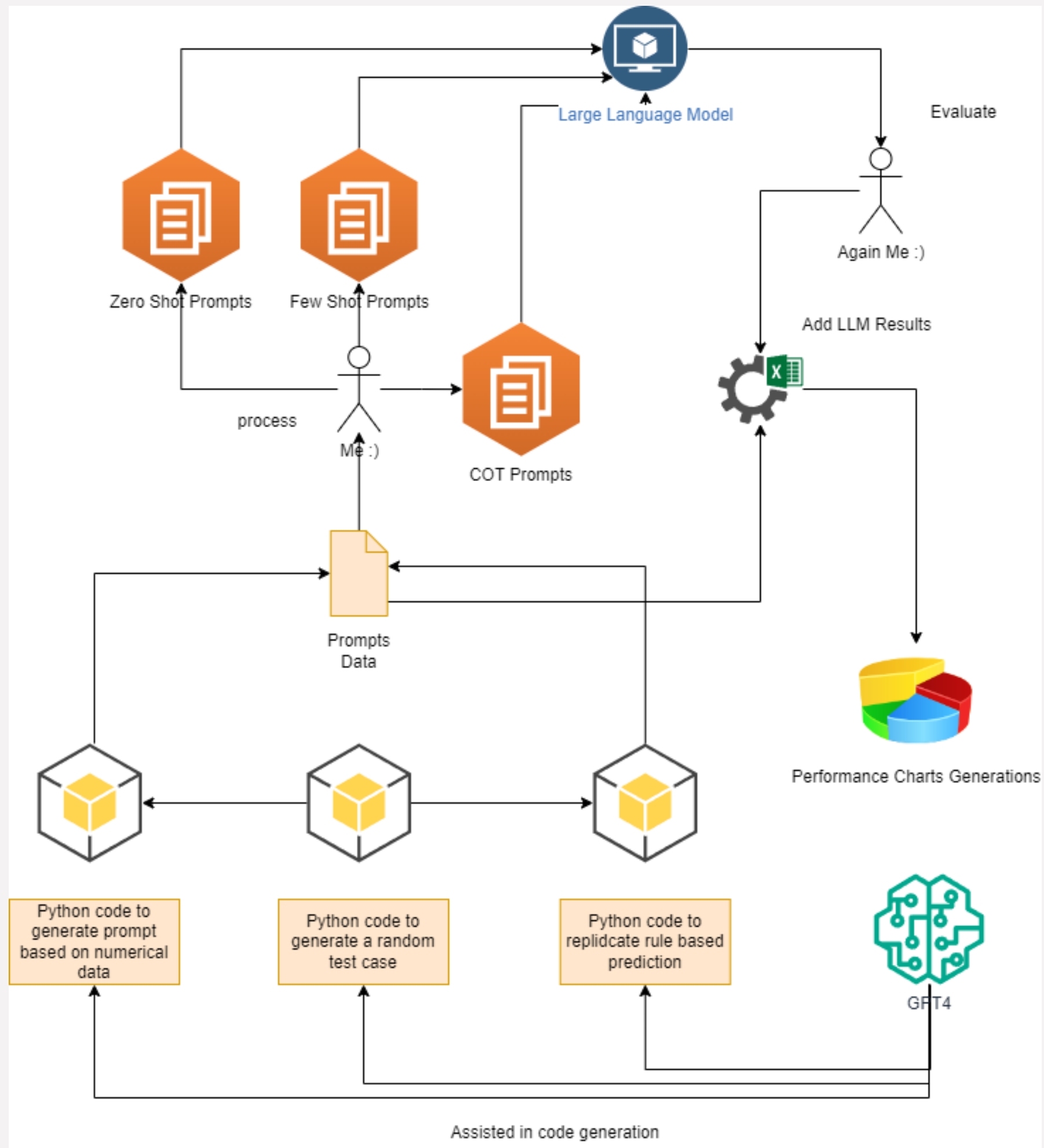
## Demo Links

1. [Zero Shot GPT4](#)
2. [Zero Shot GPT3.5](#)
3. [Few Shot GPT3.5](#)
4. [COT GPT4](#)
5. [COT GPT3.5](#)

Others models were tested locally whose chat history I plan to post in github repo at : [Link](#)



## Testing Setup





- Able to process report from Rule based project to extract required knowledge
- Converted Jess Rule Based approach, added random test case generation and correct output prediction in 2 prompts
- Helped with prompt generation for Zero shot, few shot , COT.
- Output of GPT4 can be used to COT GPT3.5 into doing better math

Capabilites of GPT4 in helping test.

1. GPT4 is a big model! If we are only considering traffic scenario a smaller model could be fine tuned and trained to work with a particular prompting method.
2. GPT3.5's performance increased with chain of thought prompting. Finetuning that makes it a good contender to be used.
3. Removing speech capabilities of the model and only focusing on predicting what we need may further reduce model size and improve
4. Integrating the model with an agent that can take action may lead to better results.

Future Improvements

Thanks!