

Actual Intelligence and Artificial Intelligence Joining Forces to Perform Statistical Analysis Turning AI Hype into Reality

Toshio Kimura Arcsine Analytics
Weiming Du Alnylam Pharmaceuticals
Songgu Xie Regeneron Pharmaceuticals

xxx 2025

03.xx.2025

AI in Drug Development

Once-in-a-Century Opportunity

The New York Times

How A.I. Is Revolutionizing Drug Development

By Steve Lohr
June 17, 2024

<https://www.nytimes.com/2024/06/17/business/ai-drugs-development-terray.html>

“

A.I. is a “once-in-a-century opportunity” for the pharmaceutical business, according to the consulting firm McKinsey & Company.

”

Actual Intelligence and Artificial Intelligence

Agenda



Introduction
LLMs can Perform
Analysis?!?!

01



The Human Workflow
Ask the right questions

03



AI²
Actual Intelligence and
Artificial Intelligence

05

02

Technology Stack
Python, SAS/R, LLM



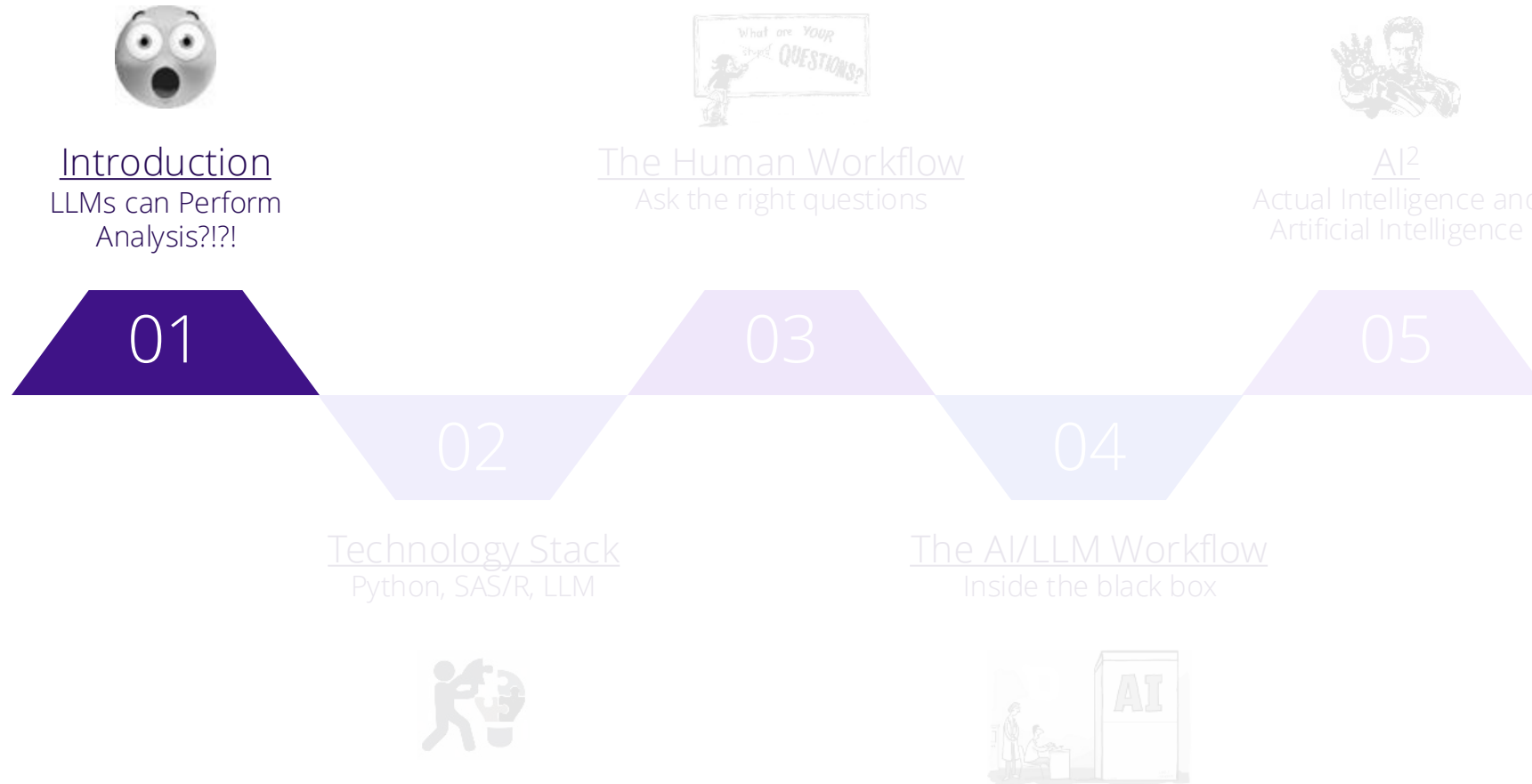
04

The AI/LLM Workflow
Inside the black box



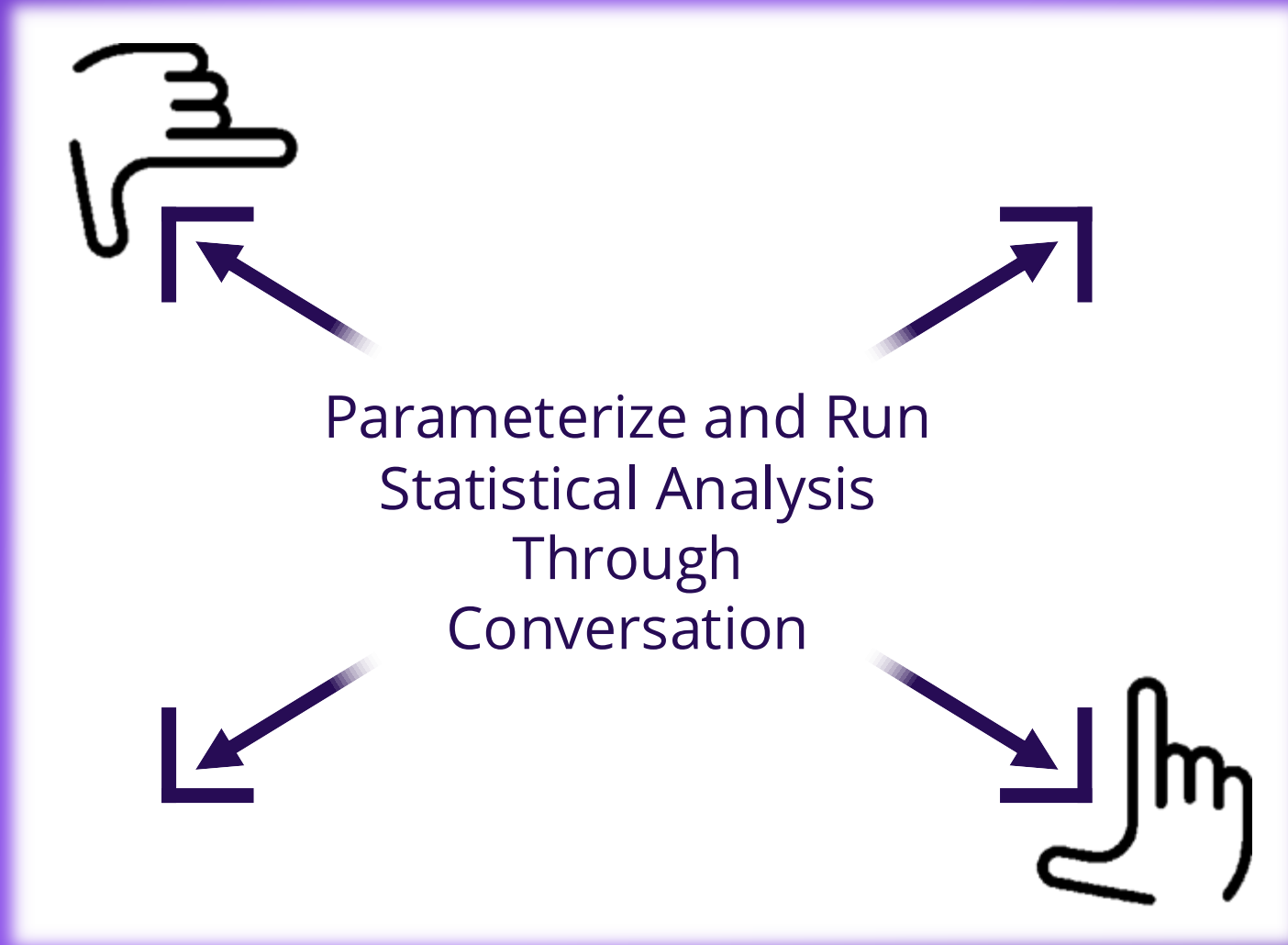
Actual Intelligence and Artificial Intelligence

Agenda



LLMs in Analysis Workflow

Scope



LLM Performing Analysis

Demonstration

Biostat Chatbot



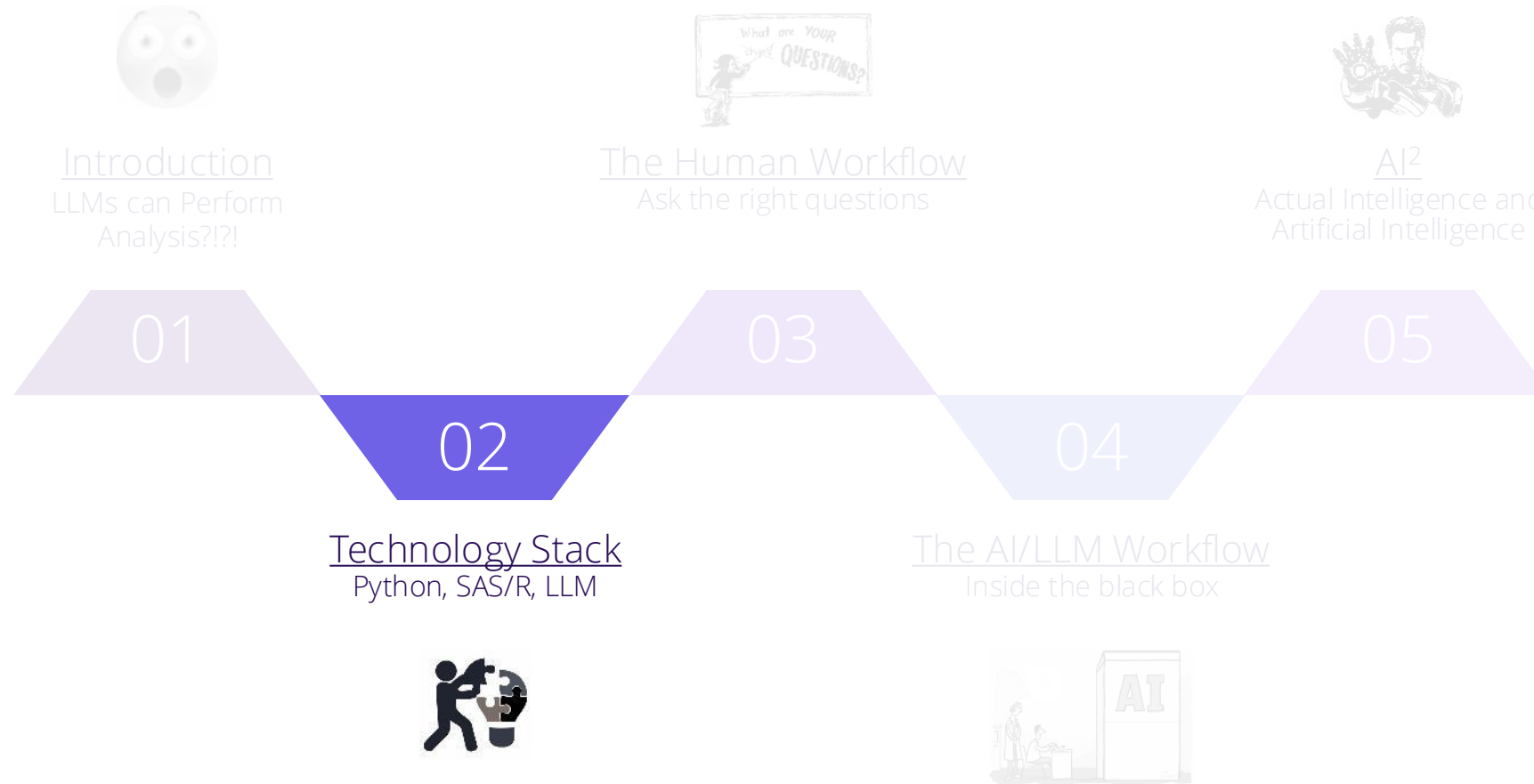
Biostat Chatbot here, what analysis do you want to perform today?

Type your message here... I

Send

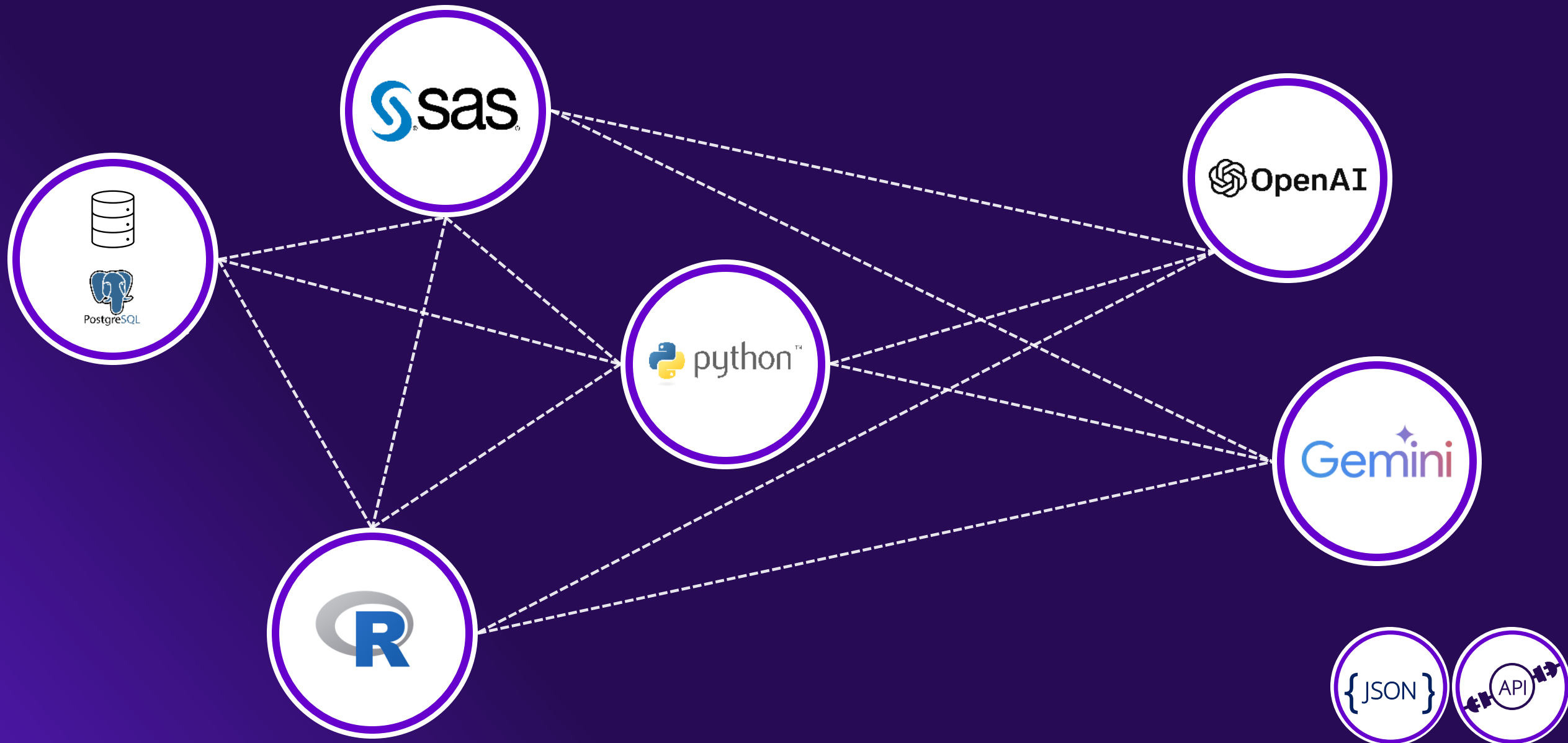
Actual Intelligence and Artificial Intelligence

Agenda



Technology Stack

Constellation of Technology to Enable the AI/LLM Analysis System



A network diagram illustrating connections between various data science and AI tools. The nodes are represented by circles containing logos or text:

- SAS**: Top center node.
- PostgreSQL**: Top left node, featuring a database cylinder icon and the PostgreSQL elephant logo.
- Python**: Center node, featuring the Python logo.
- R**: Bottom left node, featuring the R logo.
- OpenAI**: Top right node, featuring the OpenAI logo.
- Gemini**: Bottom right node, featuring the Gemini logo.
- {JSON}**: Bottom right node, featuring the text "{JSON}" in a circle.
- API**: Bottom right node, featuring the text "API" in a circle.

Dashed lines represent connections between these tools. The connections are as follows:

- SAS is connected to PostgreSQL, Python, R, OpenAI, and Gemini.
- PostgreSQL is connected to SAS and R.
- Python is connected to SAS, R, OpenAI, and Gemini.
- R is connected to SAS, Python, and Gemini.
- OpenAI is connected to SAS, Python, and Gemini.
- Gemini is connected to SAS, Python, R, and OpenAI.
- {JSON} and API are connected to Gemini.

Some of the dashed lines are labeled with "JSON" or "API", indicating the type of data exchange or integration between the tools.

Actual Intelligence and Artificial Intelligence

Agenda



Human Workflow

A Chat between Statistician and Statistical Programmer and Asking the Right Questions



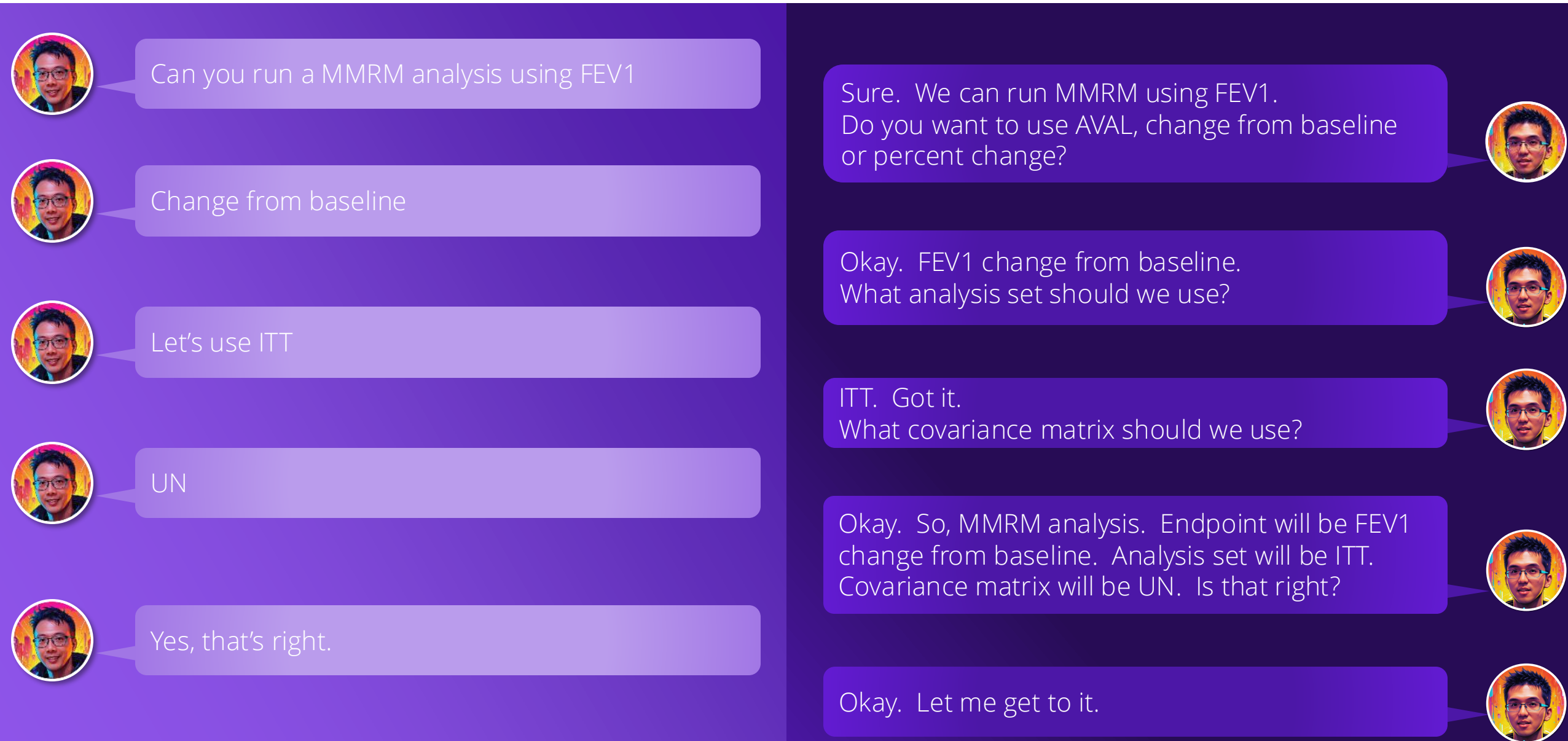
Hi, I'm Weiming.
I'm a statistician



Hi, I'm Songgu.
I'm a statistical programmer.

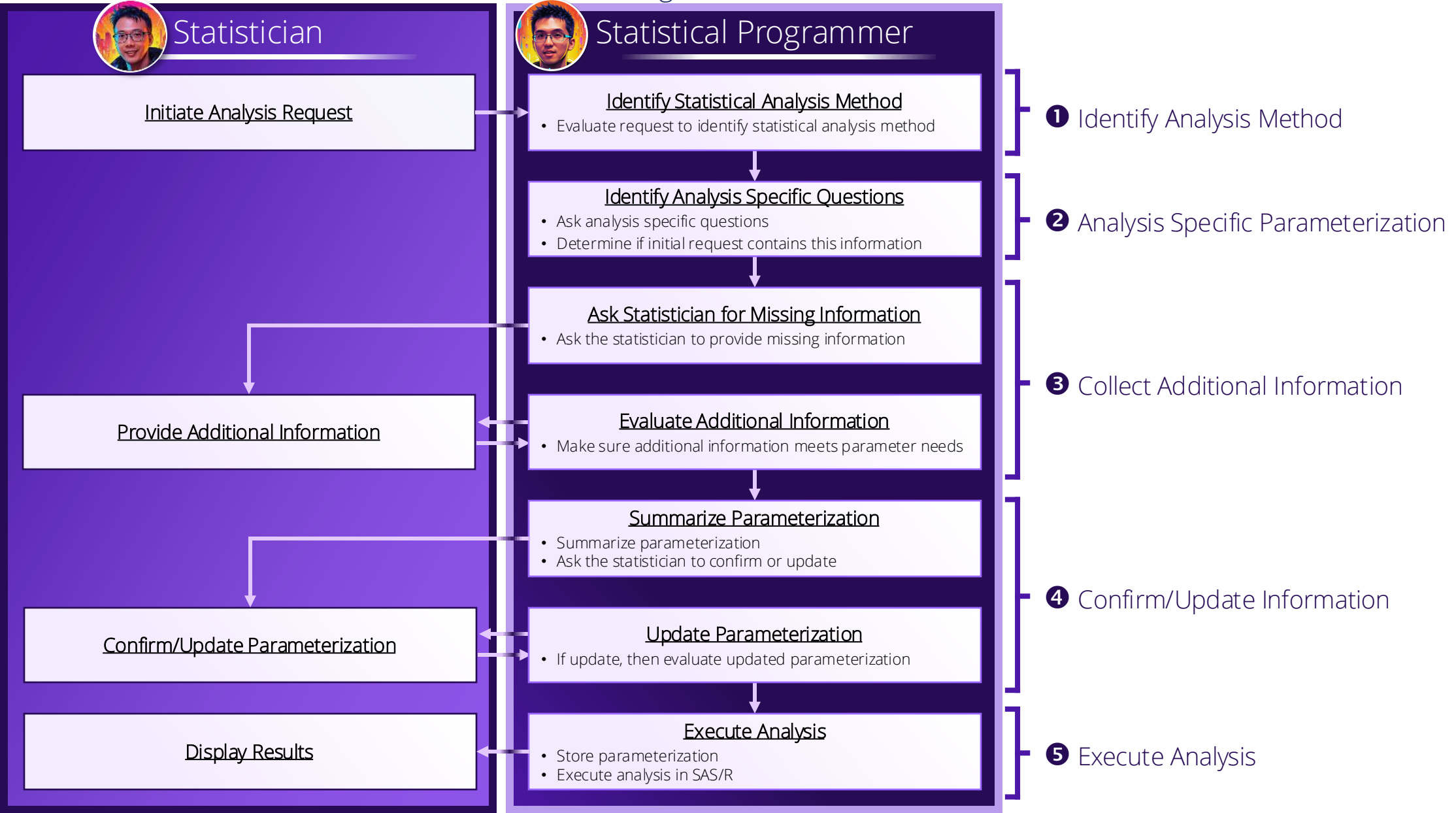
Human Workflow

A Chat between Statistician and Statistical Programmer and Asking the Right Questions



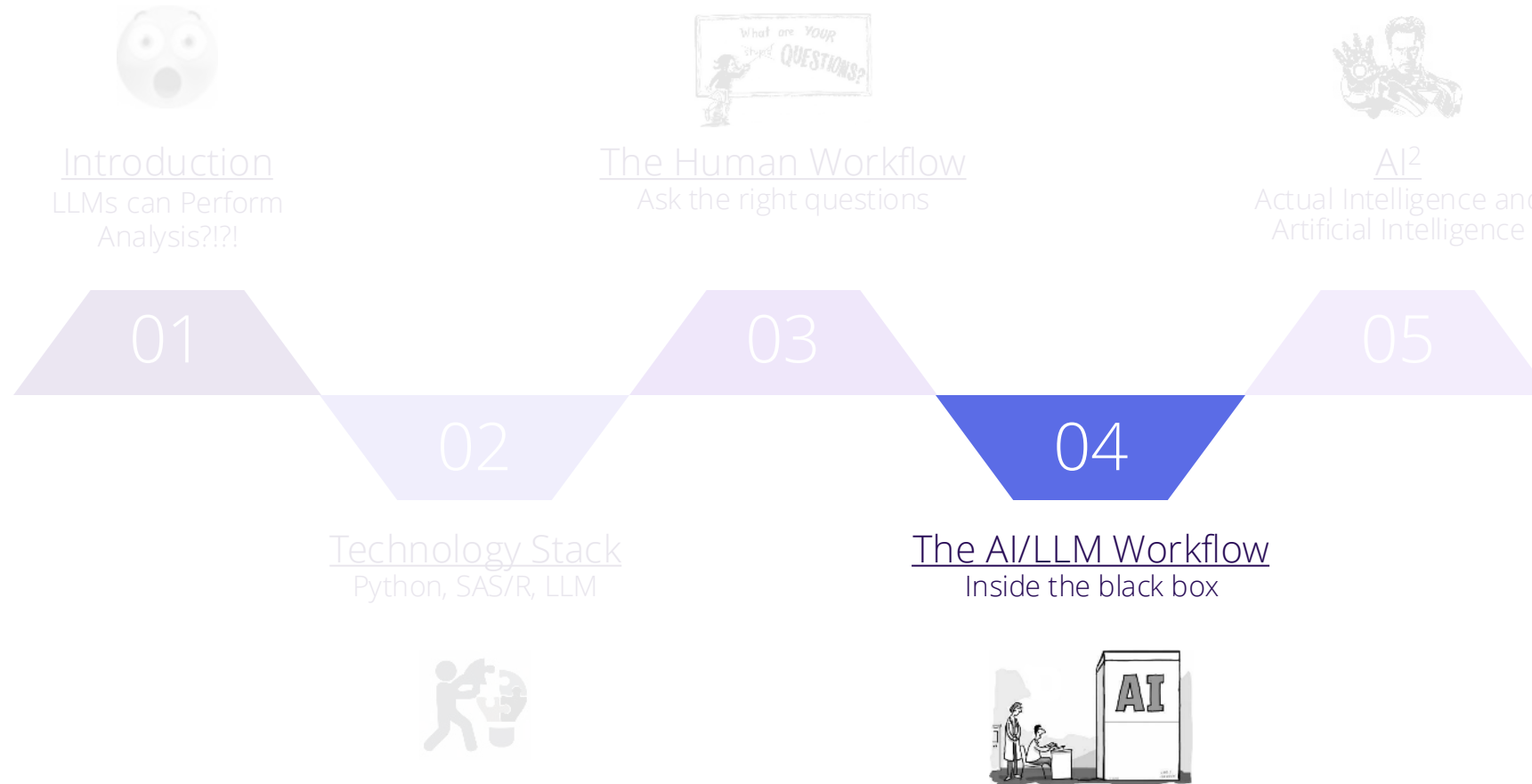
Human Workflow

Formalizing the Chat into a Structured Workflow



Actual Intelligence and Artificial Intelligence

Agenda



AI Workflow

Apply AI Magic to Generate Output



Statistician

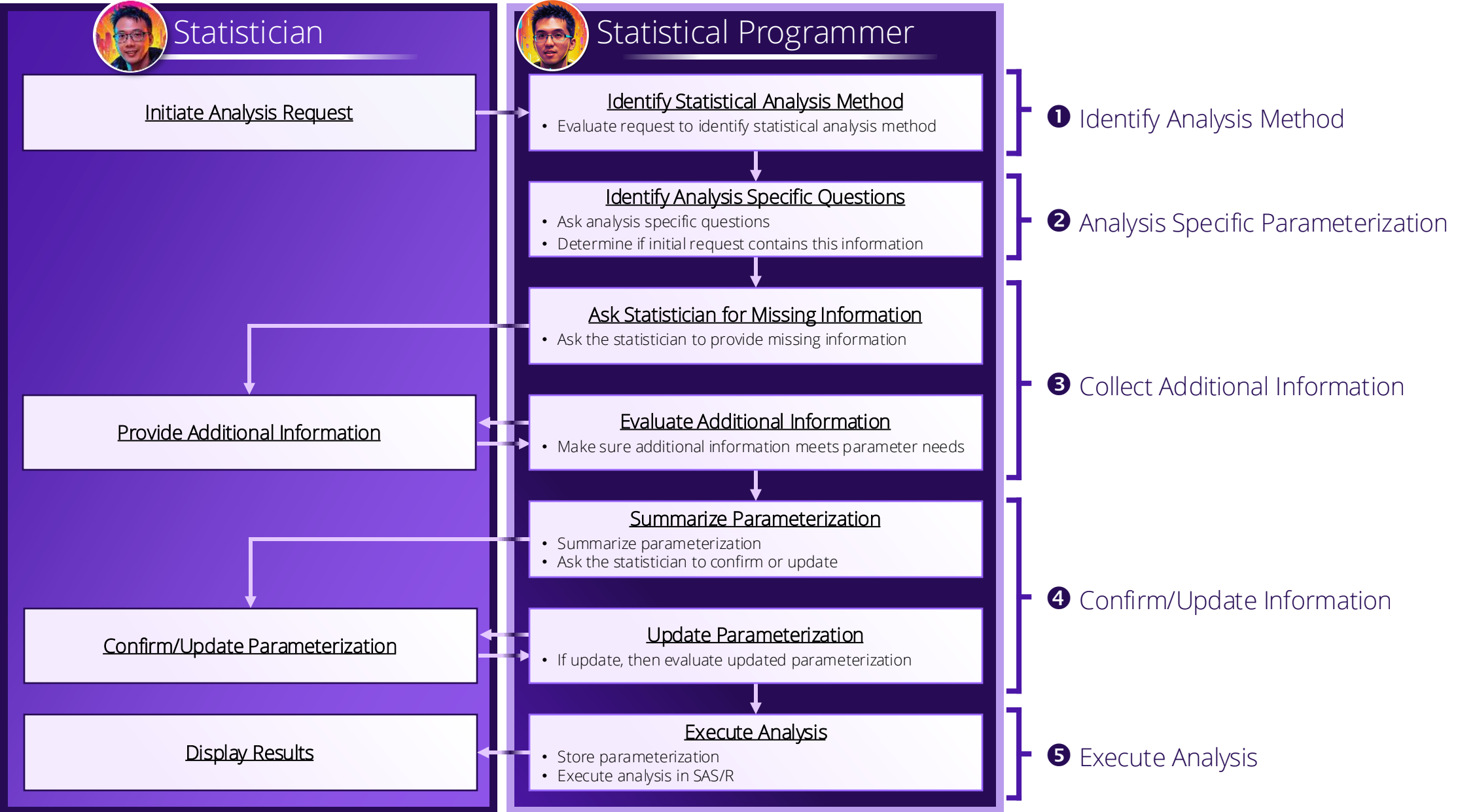
Initiate Analysis Request

Display Results



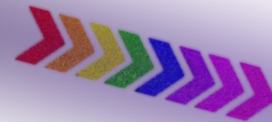
Human Workflow

The Basis for the AI Workflow



AI/LLM Workflow

Deterministic vs Probabilistic



Deterministic

Handled through traditional programming

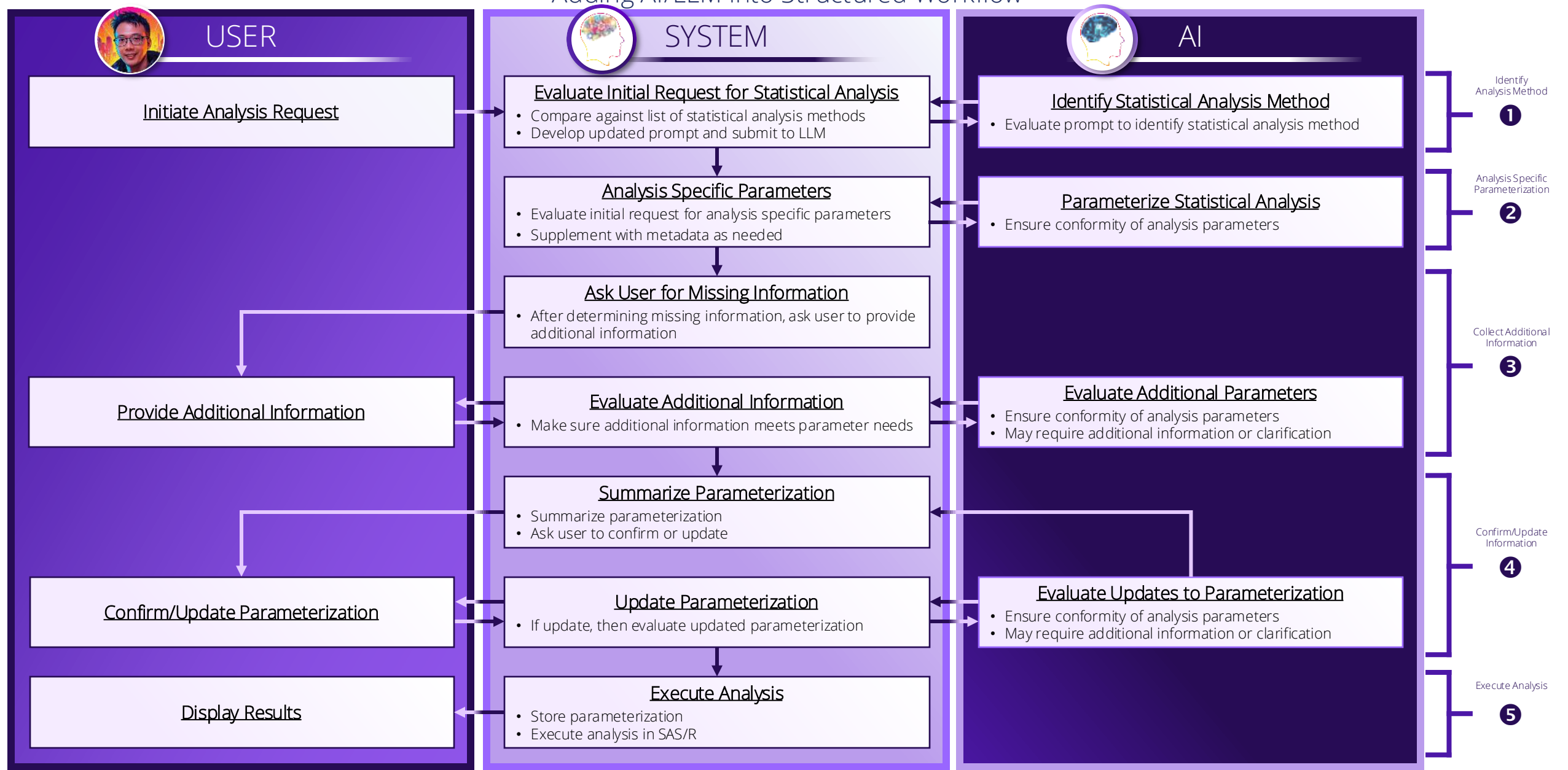


Probabilistic

Handled through AI

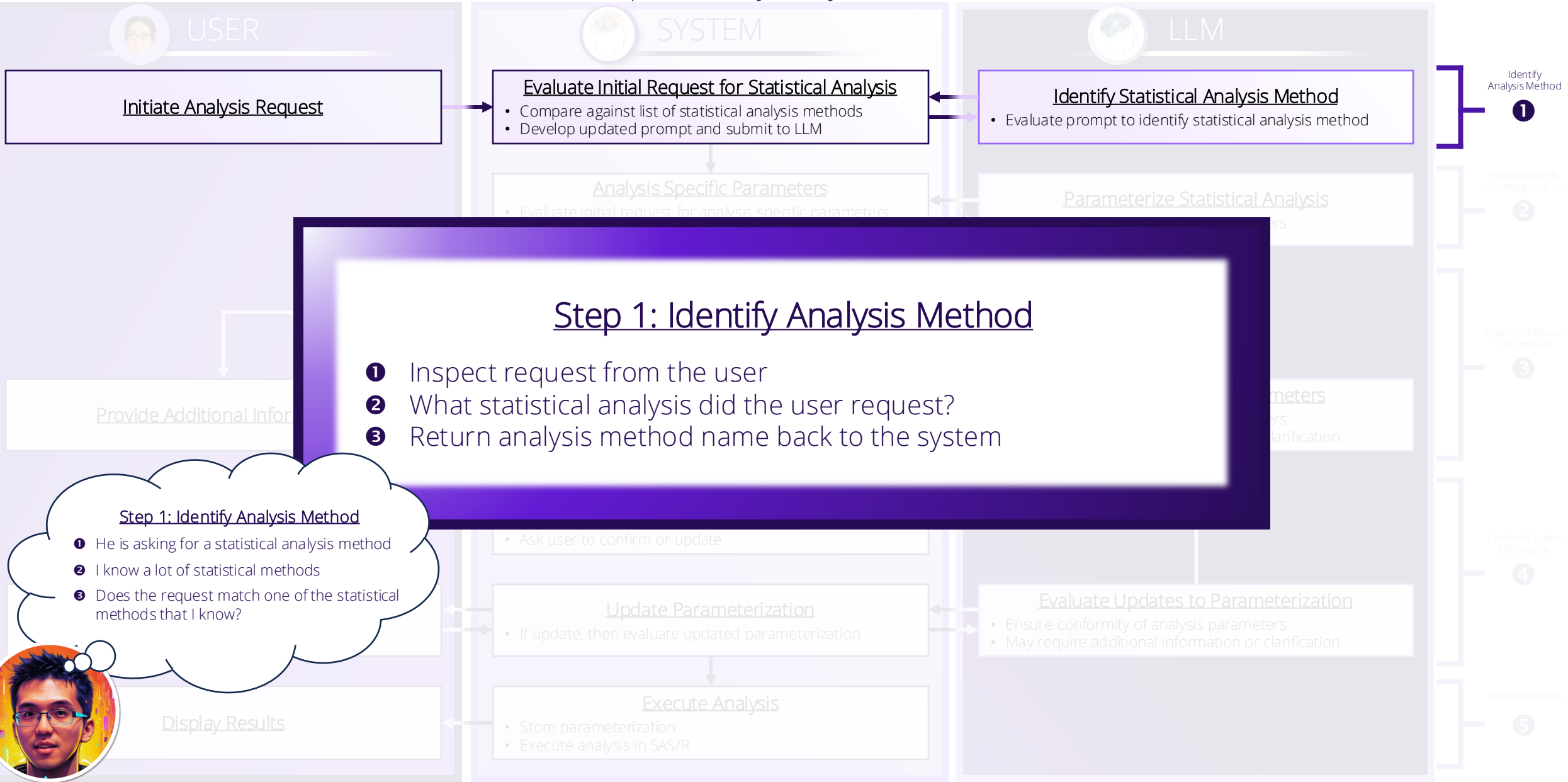
AI/LLM Structured Workflow

Adding AI/LLM into Structured Workflow



AI/LLM Structured Workflow

Step 1: Identify Analysis Method



AI/LLM Structured Workflow

Step 1: Intercept prompt and augment with additional information



I would like to run a MMRM analysis

The user requested the following:

"I would like to run a MMRM analysis"

MMRM

Here is the list of available analysis methods:

Descriptive statistics, ANCOVA, MMRM, MI ...

Which (if any) of the available analysis method is the user requesting?

Return only the name of the analysis method.

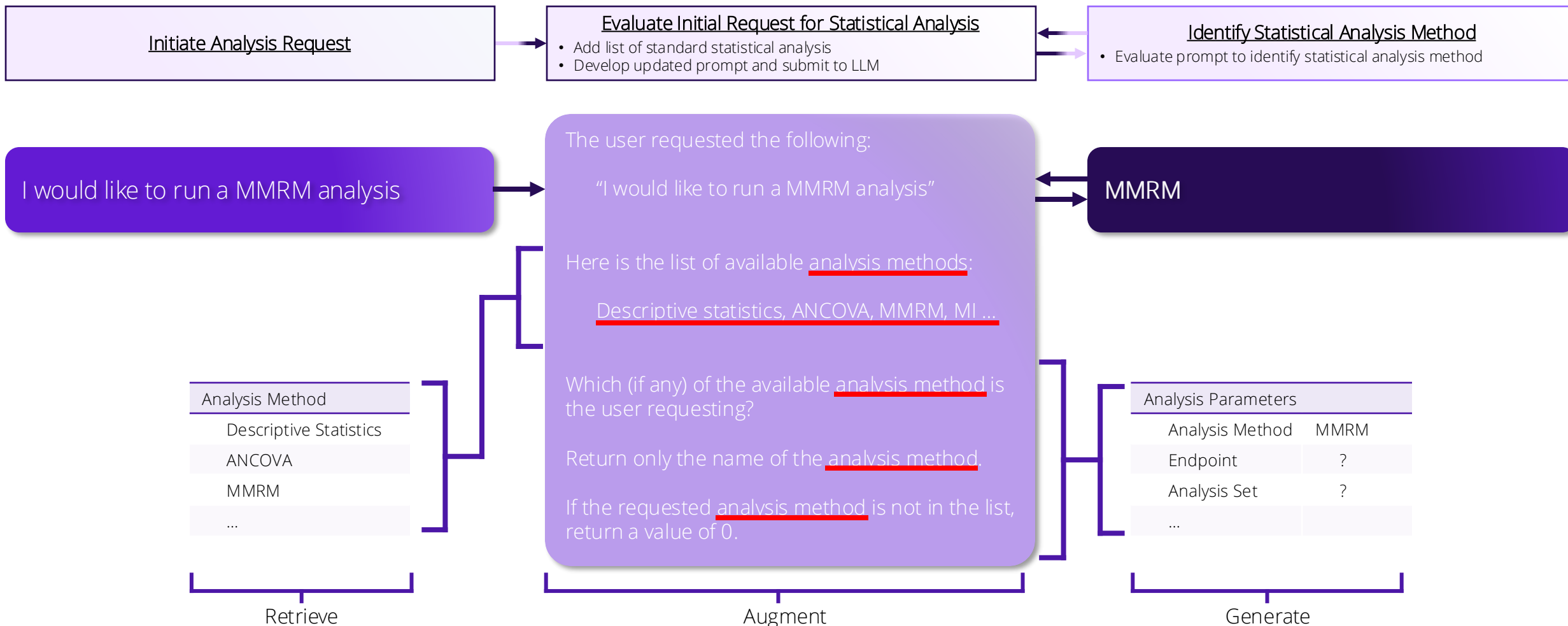
If the requested analysis method is not in the list, return a value of 0.

Step 1: Identify Analysis Method

- ❶ He is asking for a statistical analysis method
- ❷ I know a lot of statistical methods
- ❸ Does the request match one of the statistical methods that I know?



Step 1: Identify Analysis Method

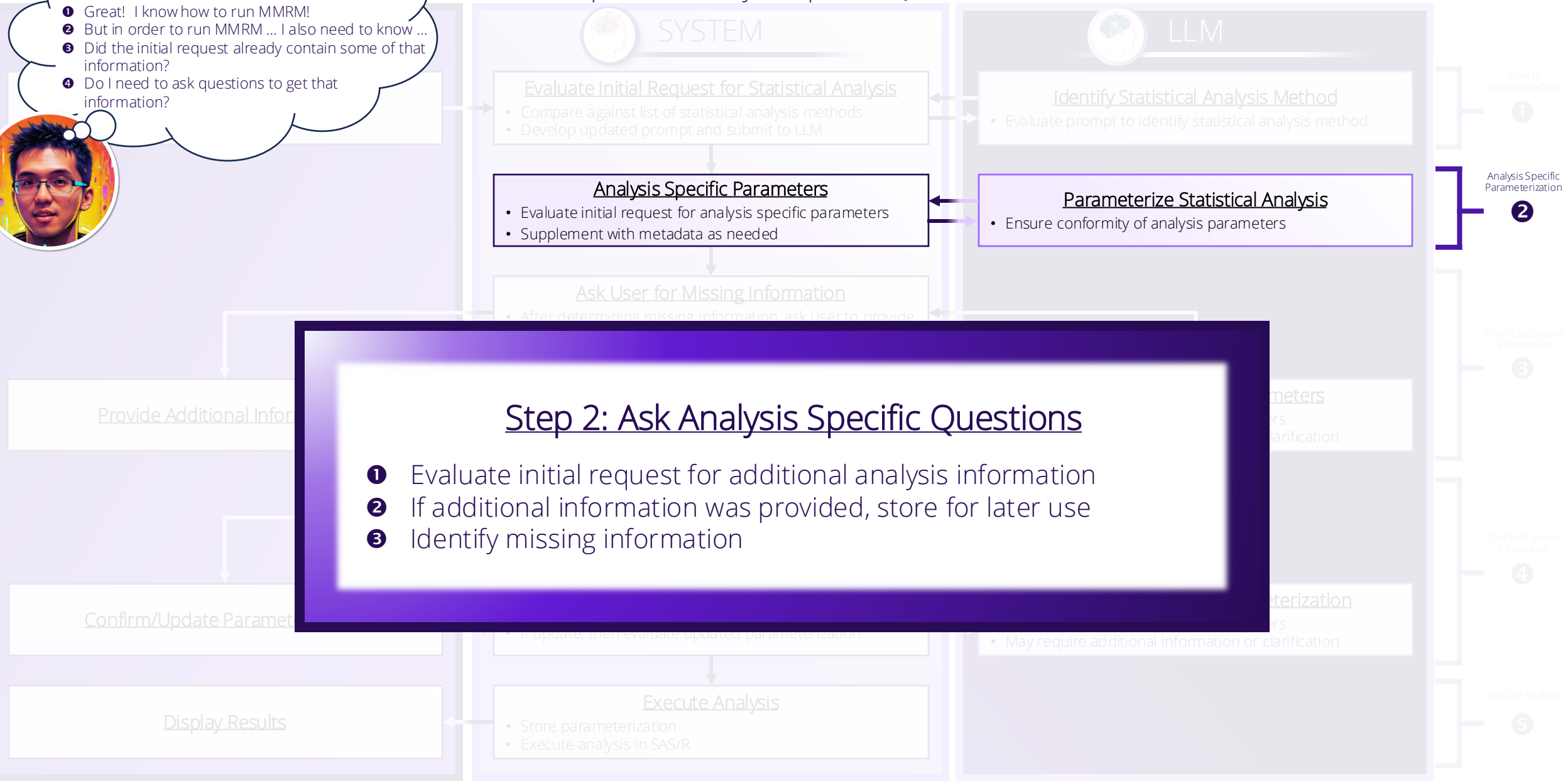


AI/LLM Structured Workflow

Step 2: Ask Analysis Specific Questions

Step 2: Analysis Specific Parameterization

- 1 Great! I know how to run MMRM!
- 2 But in order to run MMRM ... I also need to know ...
- 3 Did the initial request already contain some of that information?
- 4 Do I need to ask questions to get that information?



Identify Analysis Method

1

Analysis Specific Parameterization

2

Collect Additional Information

3

Confirm/Update Information

4

Execute Analysis

5

AI/LLM Structured Workflow

Step 2: Intercept prompt and augment with additional information



Analysis Parameters		
1 ✓	Analysis Method	MMRM
2	Ask about next → Endpoint	?
	Covariance Matrix	?
	...	

Step 2: Analysis Specific Parameterization

- 1 Great! I know how to run MMRM!
- 2 But in order to run MMRM ... I also need to know ...
- 3 Did the initial request already contain some of that information?
- 4 Do I need to ask questions to get that information?

The user requested the following:

"I would like to run a MMRM analysis."

Here is the list of available endpoints:

FEV1, FEV1PRED, FVC ...

Which (if any) of the available endpoints is the user requesting?

Return only the PARAMCD value for the endpoints.

If the requested endpoints is not in the list, return a value of 0.

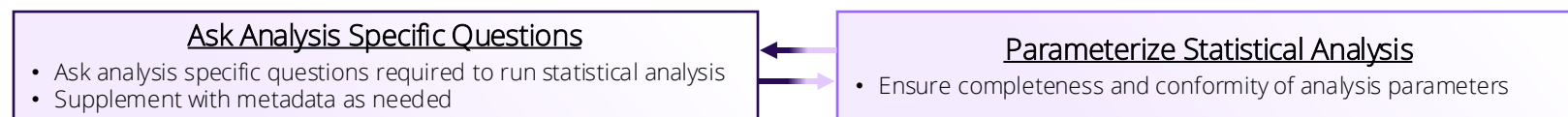
0

4

3



Step 2: Ask Analysis Specific Questions



Analysis Parameters		
✓	Analysis Method	MMRM
→	Endpoint	?
	Covariance Matrix	?
	...	

Endpoints	
FEV1	
FEV1PRED	
FVC	
...	

Retrieve

The user requested the following:

"I would like to run a MMRM analysis."

Here is the list of available endpoints:

FEV1, FEV1PRED, FVC ...

Which (if any) of the available endpoints is the user requesting?

Return only the PARAMCD value for the endpoints.

If the requested endpoints is not in the list, return a value of 0.

Augment

0

Analysis Parameters		
	Analysis Method	MMRM
	Endpoint	0
	Covariance Matrix	?
	...	

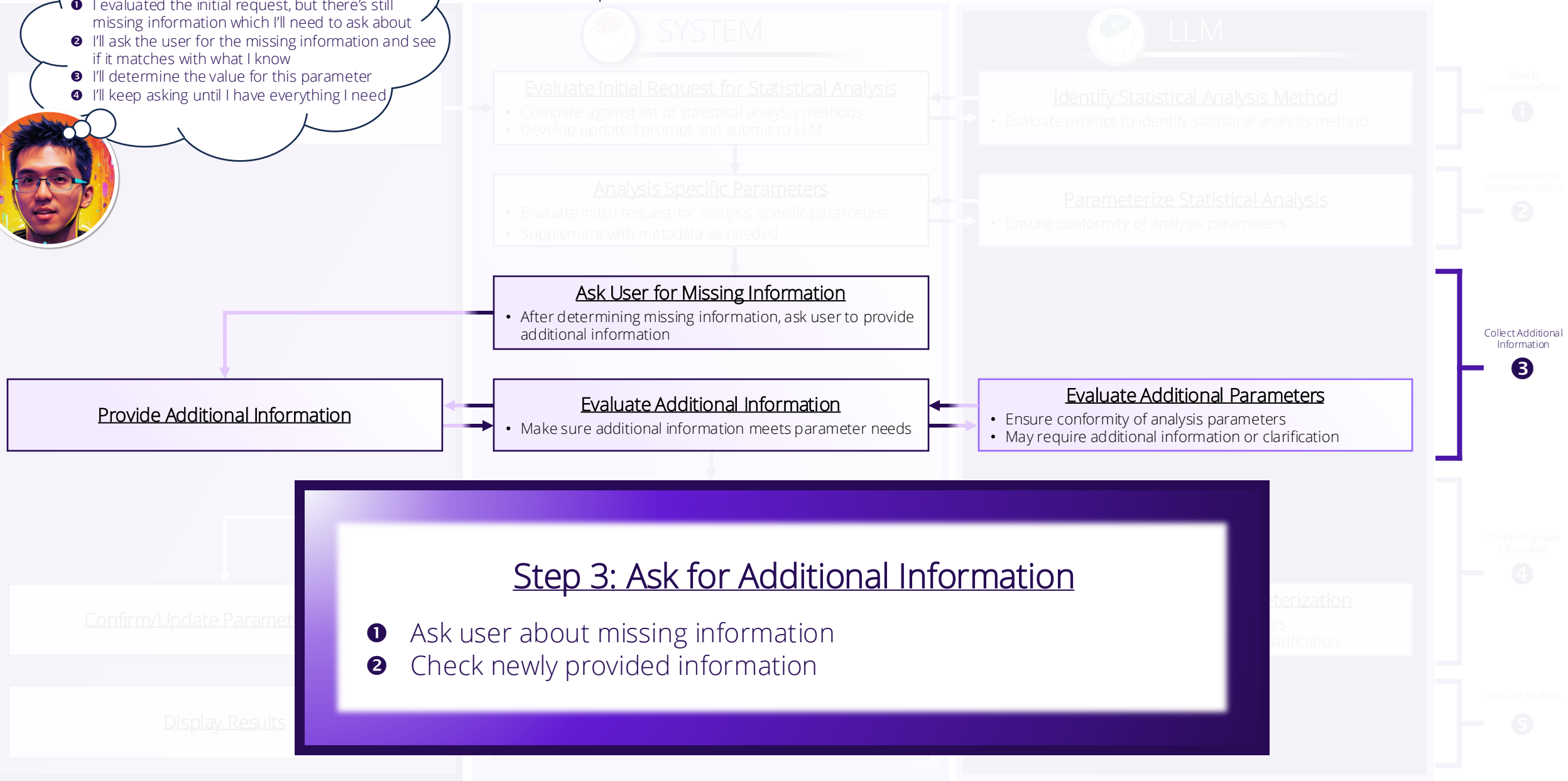
Generate

AI/LLM Structured Workflow

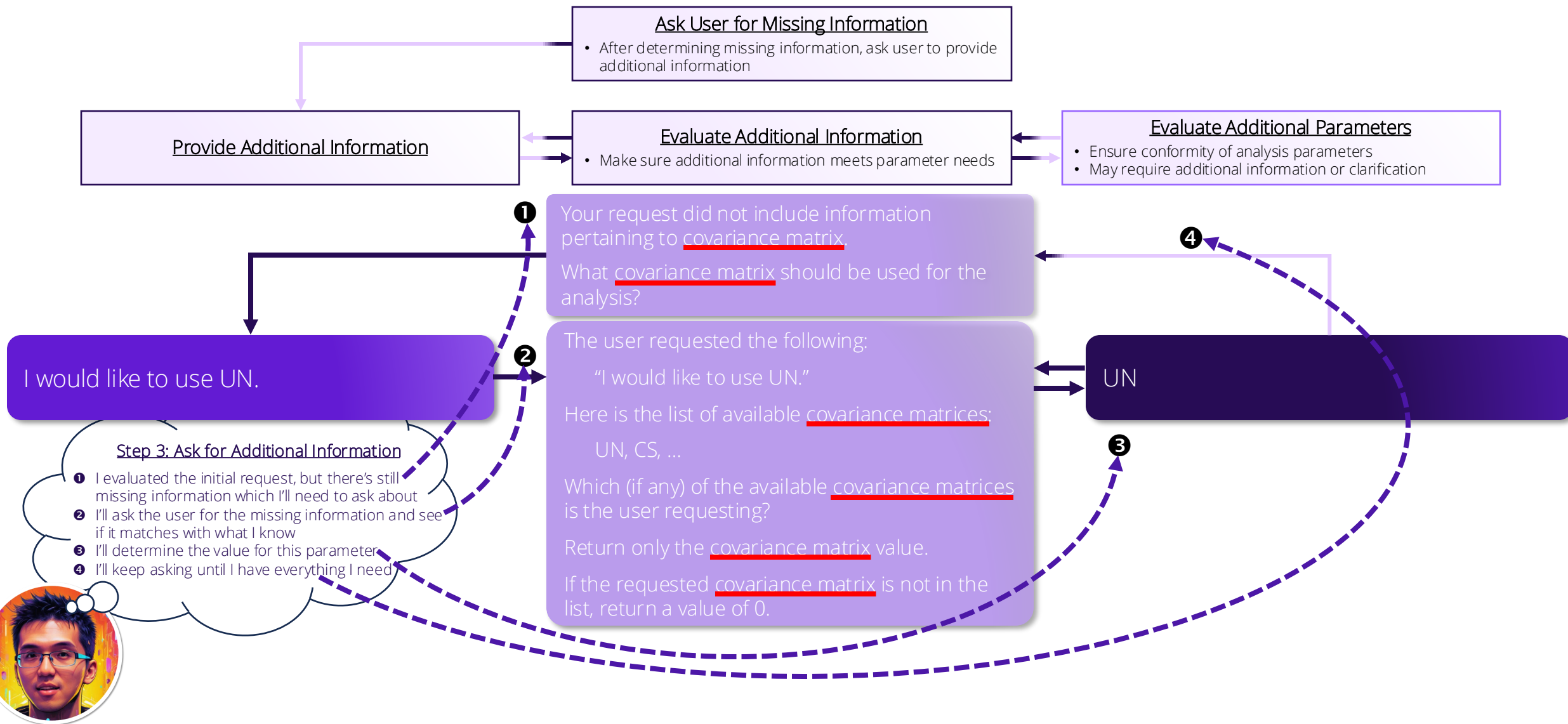
Step 3: Ask for Additional Information

Step 3: Ask for Additional Information

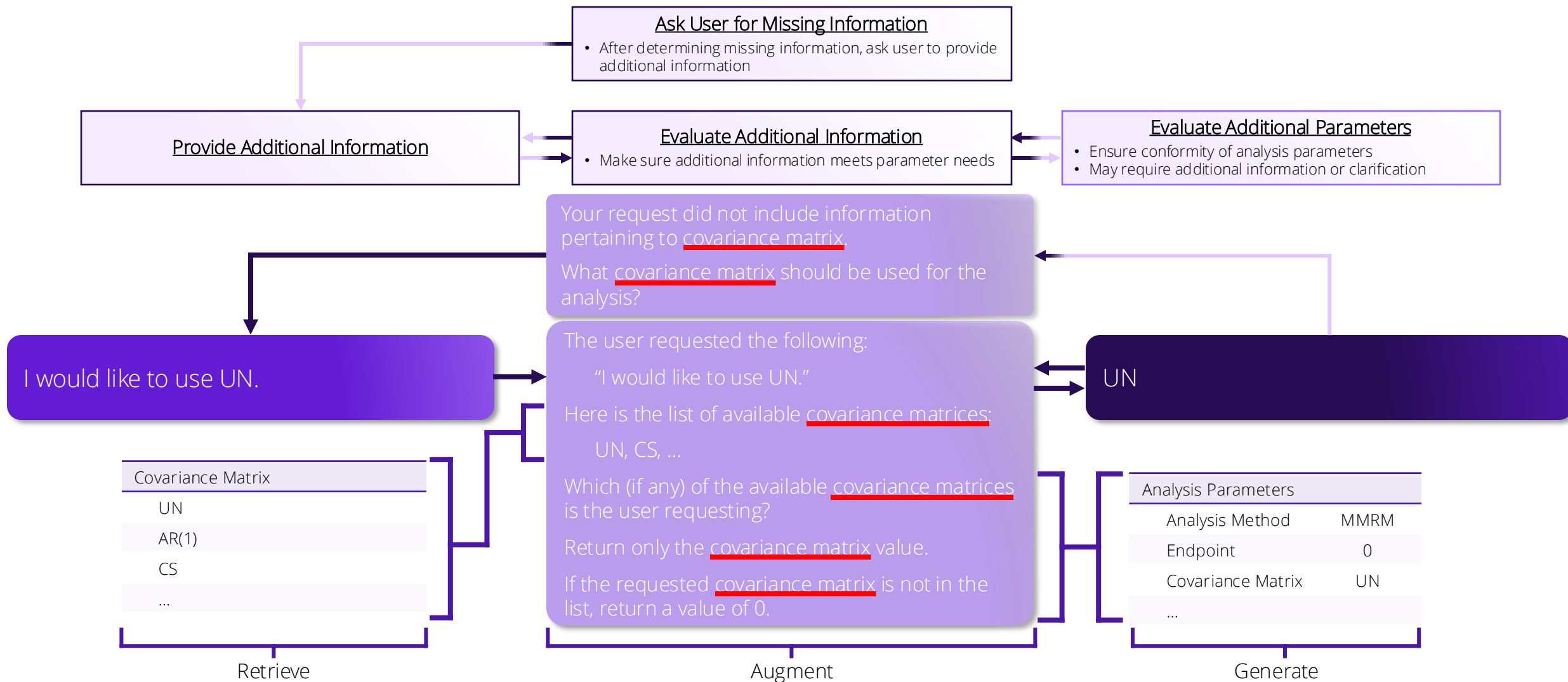
- 1 I evaluated the initial request, but there's still missing information which I'll need to ask about
- 2 I'll ask the user for the missing information and see if it matches with what I know
- 3 I'll determine the value for this parameter
- 4 I'll keep asking until I have everything I need



Step 3: Ask for Additional Information



Step 3: Ask for Additional Information



AI/LLM Structured Workflow

Step 4: Confirm/Update Information

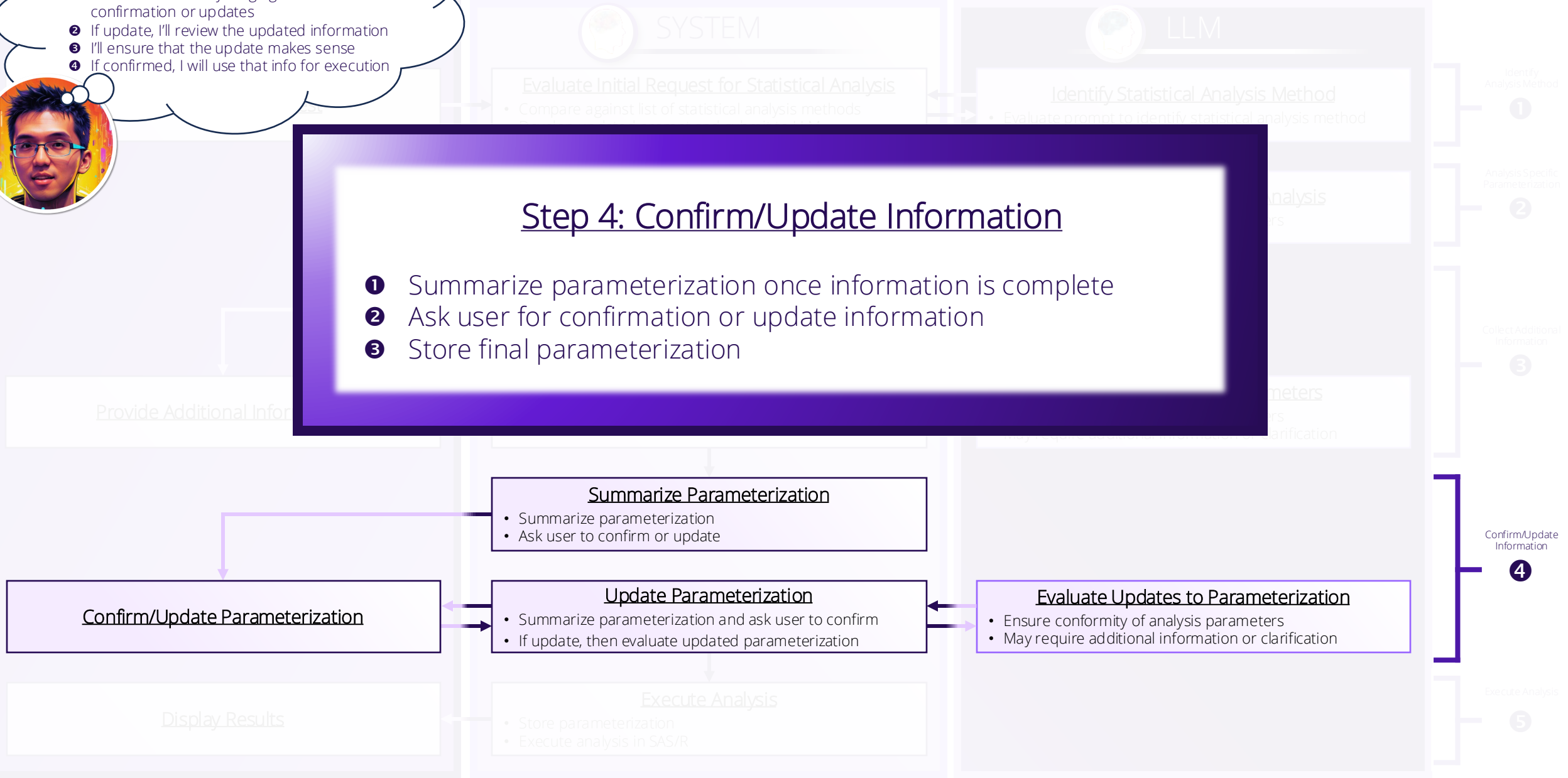
Step 4: Confirm/Update Information

- 1 I'll summarize everything I gathered and ask for confirmation or updates
- 2 If update, I'll review the updated information
- 3 I'll ensure that the update makes sense
- 4 If confirmed, I will use that info for execution



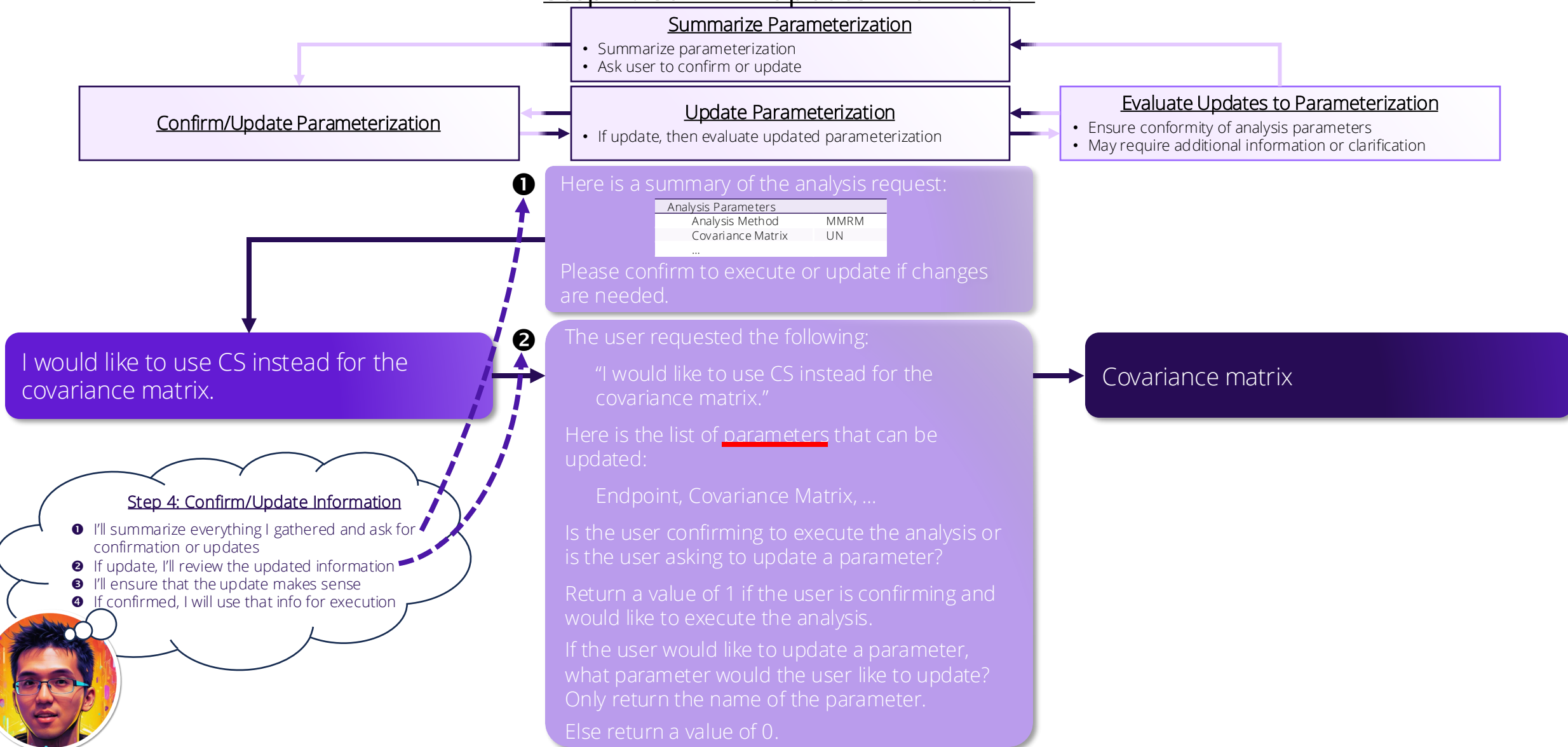
Step 4: Confirm/Update Information

- 1 Summarize parameterization once information is complete
- 2 Ask user for confirmation or update information
- 3 Store final parameterization



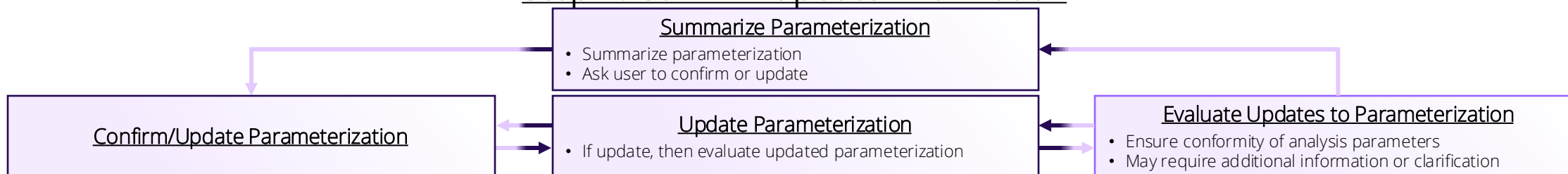
Step 4: Intercept prompt and augment with additional information

Step 4: Confirm/Update Information



Step 4: Intercept prompt and augment with additional information

Step 4: Confirm/Update Information



3

The user requested the following:
 "I would like to use CS instead for the covariance matrix."
 Here is the list of available covariance matrices:
 UN, CS, ...
 Which (if any) of the available covariance matrices is the user requesting?
 Return only the covariance matrix value.
 If the requested covariance matrix is not in the list, return a value of 0.

Covariance matrix

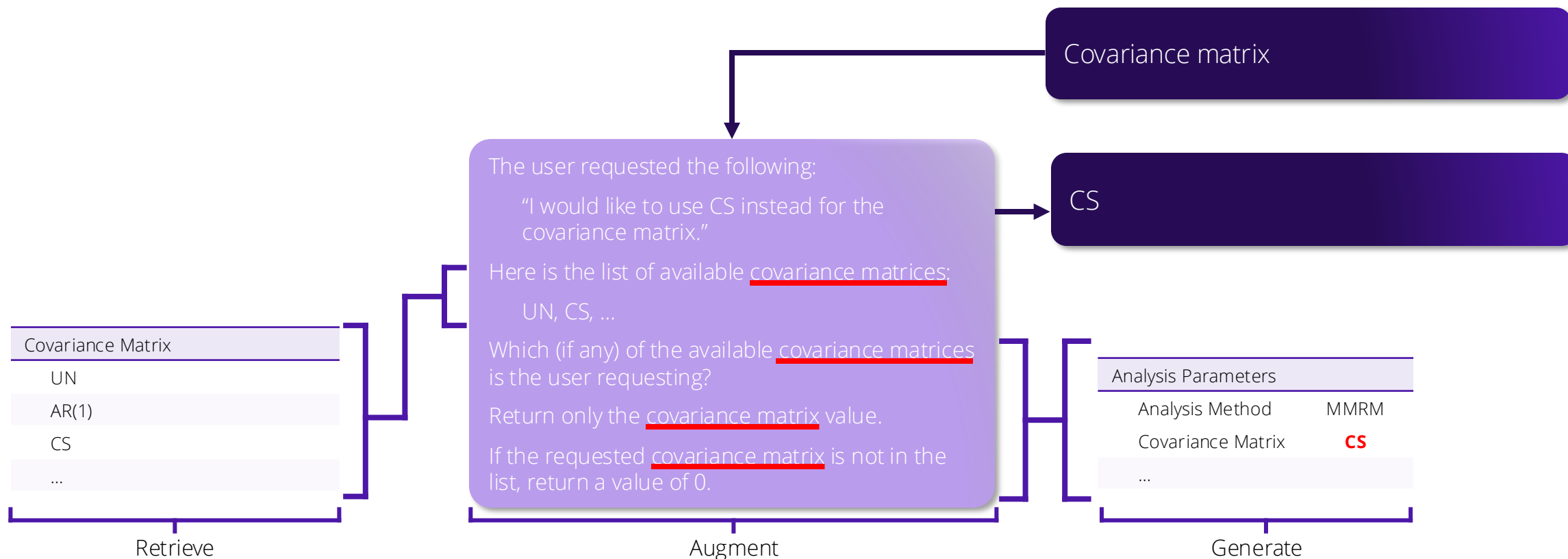
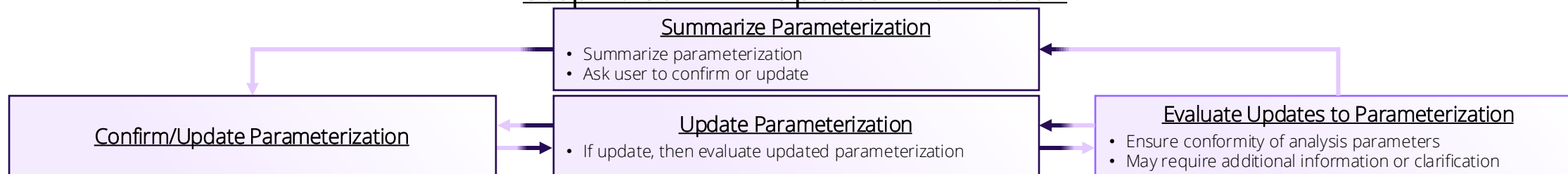
CS

Step 4: Confirm/Update Information

- 1 I'll summarize everything I gathered and ask for confirmation or updates
- 2 If update, I'll review the updated information
- 3 I'll ensure that the update makes sense
- 4 If confirmed, I will use that info for execution

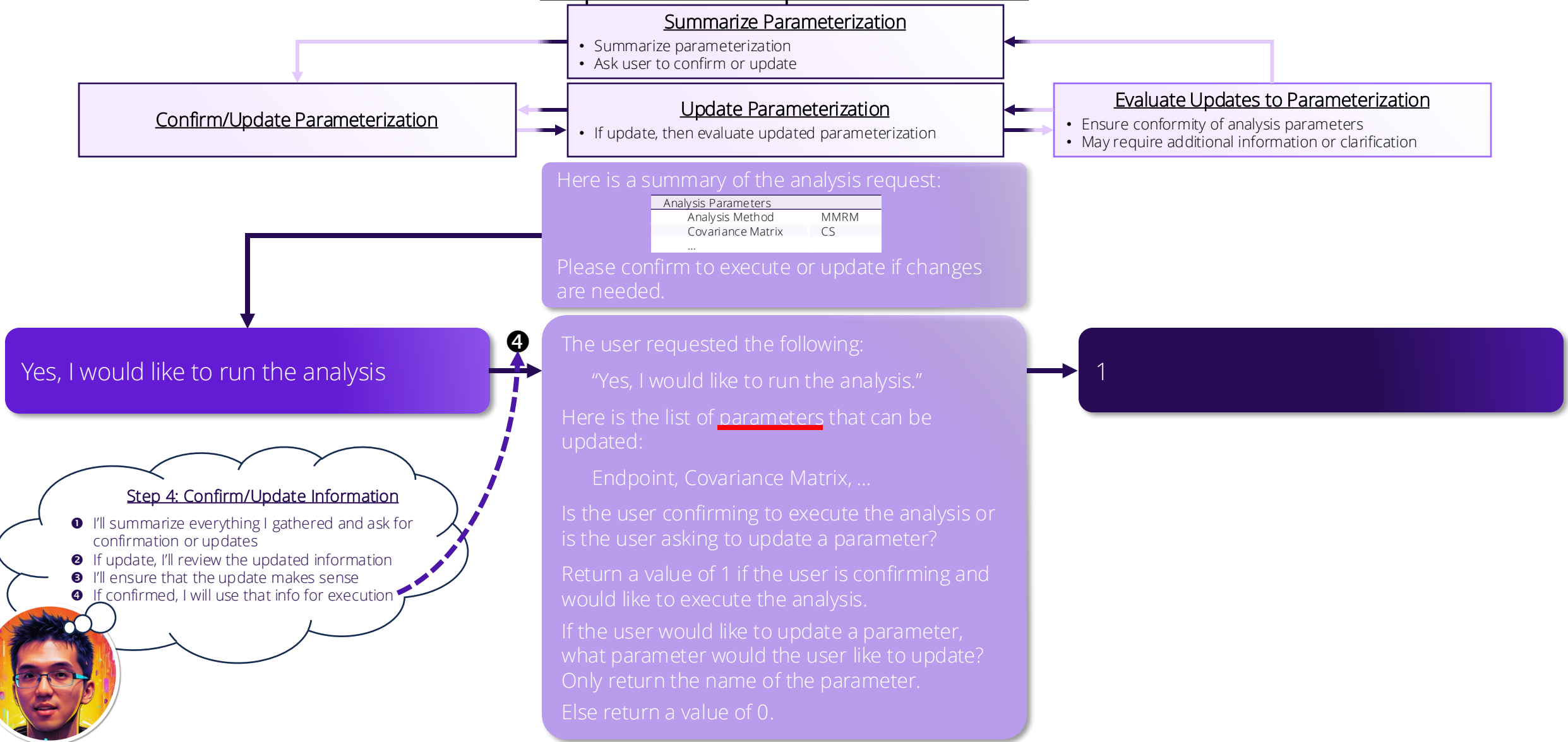


Step 4: Confirm/Update Information



Step 4: Intercept prompt and augment with additional information

Step 4: Confirm/Update Information

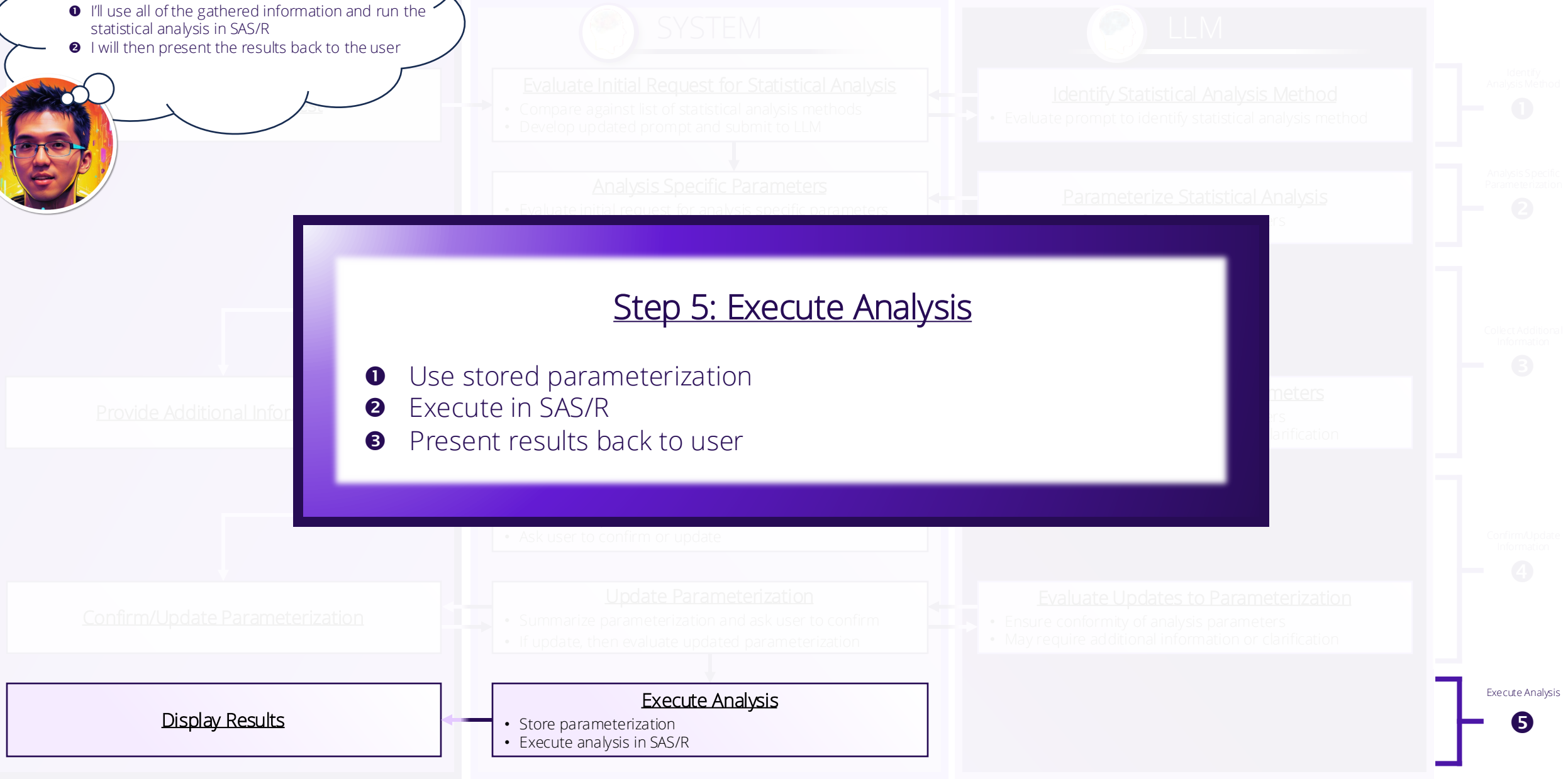
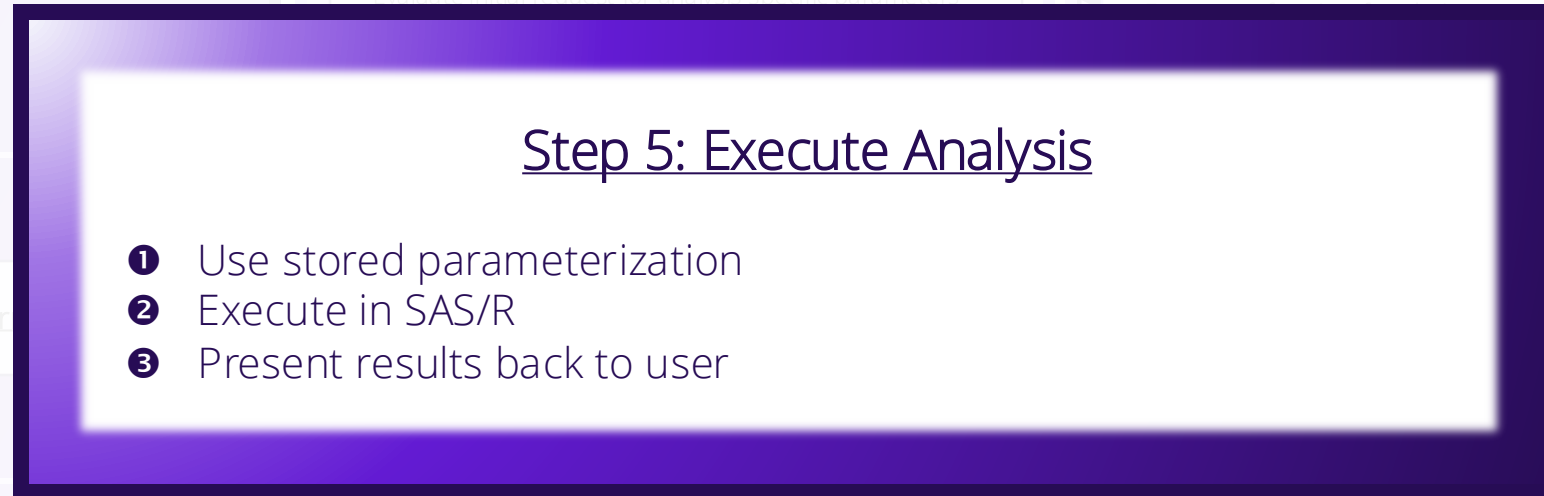


AI/LLM Structured Workflow

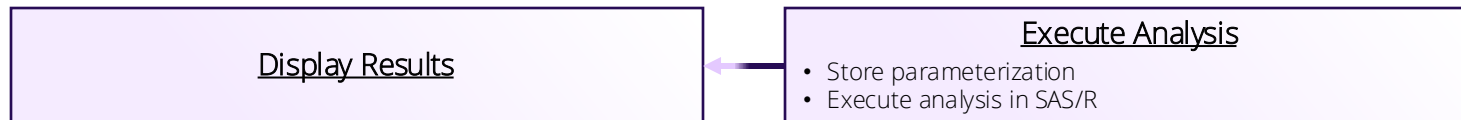
Step 5: Execute Analysis

Step 5: Execute Analysis

- ❶ I'll use all of the gathered information and run the statistical analysis in SAS/R
- ❷ I will then present the results back to the user



Step 5: Execute Analysis



Here are the results:

2

		Statistics based on ANCOVA										
		Summary Statistics				LS Means			Covariate Effect: Difference vs Reference			
Visit	Treatment	N	n	Mean	SE	Est	SE	95% CI	Est	SE	95% CI	P-value
Baseline	PBO (reference)	xxx	xxx	xxx	xx							
	Trt 1	xxx	xxx	xxx	xx							
	Trt 2	xxx	xxx	xxx	xx							
Week 4	PBO (reference)	xxx	xxx	xxx	xx	xxx	x,x	xxx,xxx				
	Trt 1	xxx	xxx	xxx	xx	xxx	x,x	xxx,xxx	xxx	x,x	xxx,xxx	0.0000
	Trt 2	xxx	xxx	xxx	xx	xxx	x,x	xxx,xxx	xxx	x,x	xxx,xxx	0.0000

The following information will be sent to SAS for analysis:

Analysis Parameters	
Analysis Method	MMRM
Covariance Matrix	CS
...	

The analysis has been executed using SAS.

		Statistics based on ANCOVA										
		Summary Statistics			LS Means			Covariate Effect: Difference vs Reference				
Visit	Treatment	N	n	Mean	SE	Est	SE	95% CI	Est	SE	95% CI	P-value
Baseline	PBO (reference)	xxx	xxx	xxx	x,x							
	Trt 1	xxx	xxx	xxx	x,x							
	Trt 2	xxx	xxx	xxx	x,x							
Week 4	PBO (reference)	xxx	xxx	xxx	x,x	xxx	x,x	xxx, xxx				
	Trt 1	xxx	xxx	xxx	x,x	xxx	x,x	xxx, xxx	xxx	x,x	xxx, xxx	0.0000
	Trt 2	xxx	xxx	xxx	x,x	xxx	x,x	xxx, xxx	xxx	x,x	xxx, xxx	0.0000



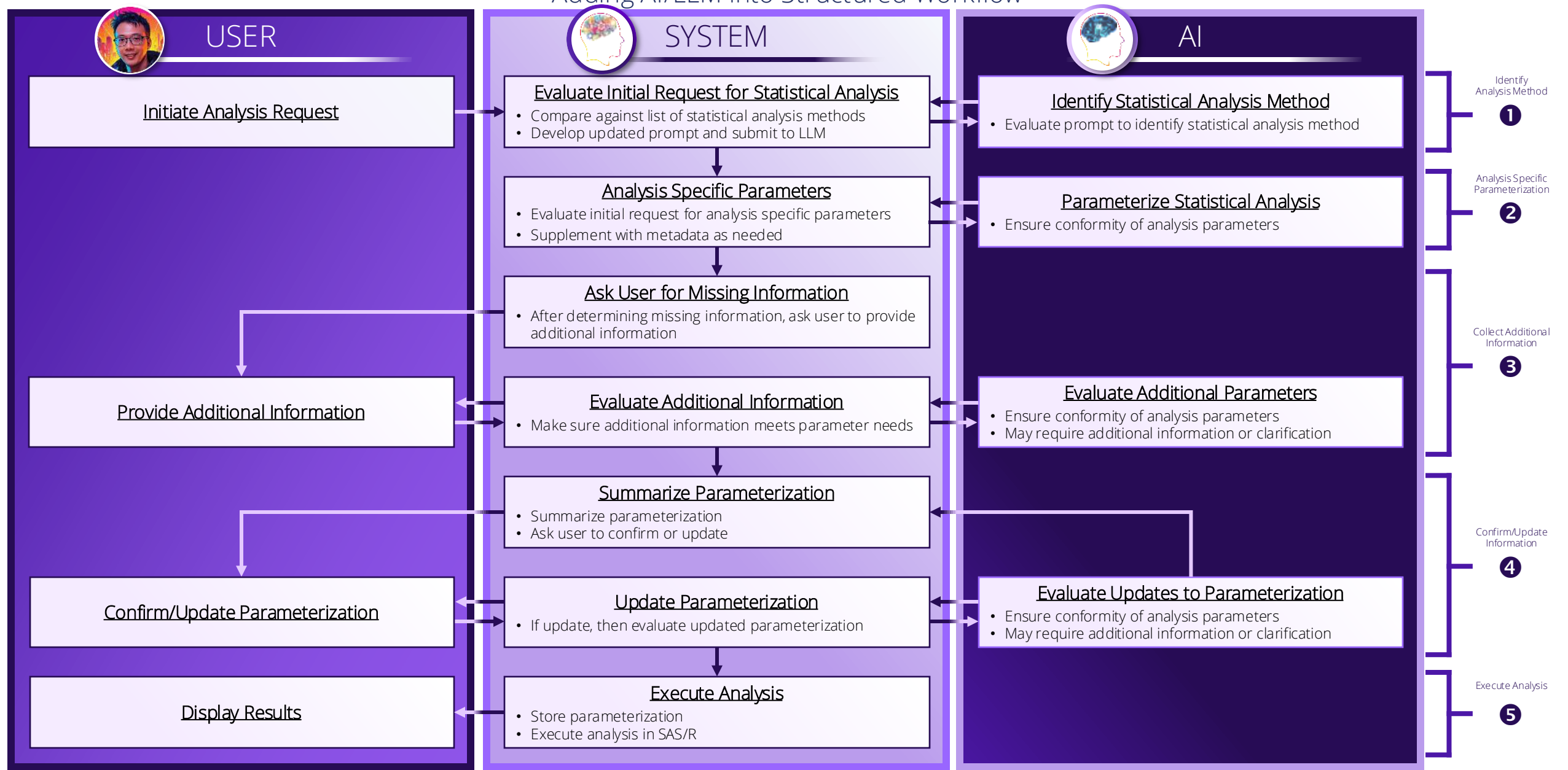
Step 5: Execute Analysis

- 1 I'll use all of the gathered information and run the statistical analysis in SAS/R
- 2 I will then present the results back to the user



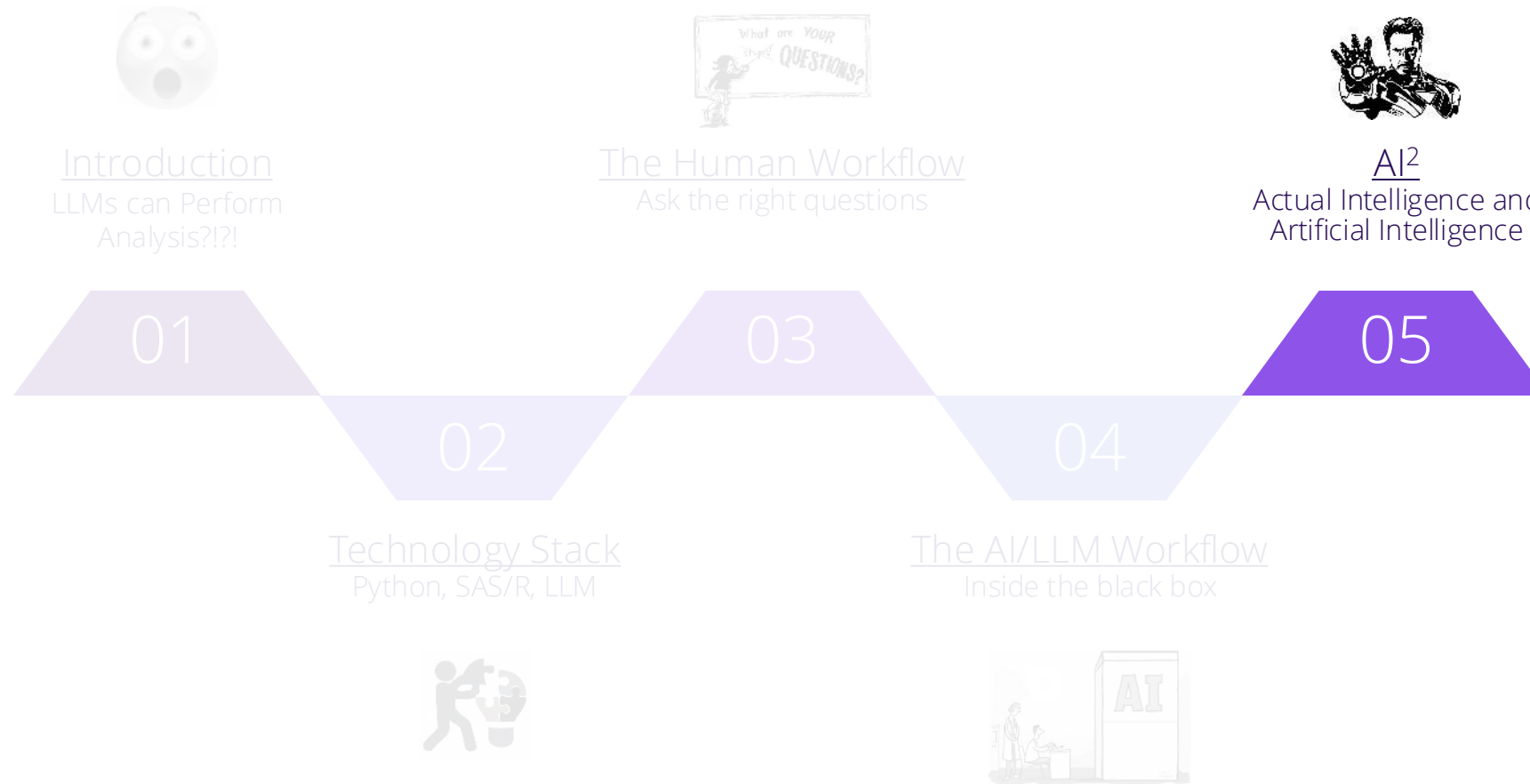
AI/LLM Structured Workflow

Adding AI/LLM into Structured Workflow



Actual Intelligence and Artificial Intelligence

Agenda



Actual Intelligence and Artificial Intelligence

Joining Forces to Perform Statistical Analysis



Actual Intelligence and Artificial Intelligence

Joining Forces to Perform Statistical Analysis



Actual Intelligence and Artificial Intelligence

“AI will cure diseases”

THE WALL STREET JOURNAL.

At Moderna, OpenAI's GPTs Are
Changing Almost Everything

By Isabelle Bousquette, April 24, 2024

<https://www.wsj.com/articles/at-moderna-openais-gpts-are-changing-almost-everything-6ff4c4a5>

“

‘People literally talk about how AI is going to cure diseases someday, and I think this is a very meaningful first step,’ said OpenAI CEO Sam Altman

”

Actual Intelligence and Artificial Intelligence

"People using AI will cure diseases"



“

People using AI will cure diseases.

”

Actual Intelligence and Artificial Intelligence

"We, using AI, will cure diseases"



“

We, using AI, will cure diseases.

”

Actual Intelligence and Artificial Intelligence

Agenda



Introduction
LLMs can Perform
Analysis?!?!
01



The Human Workflow
Ask the right questions
03



AI²
Actual Intelligence and
Artificial Intelligence
05

toshio.kimura@arcsineanalytics.com

www.arcsineanalytics.com

02

Technology Stack
Python, SAS/R, LLM



04

The AI/LLM Workflow
Inside the black box

