

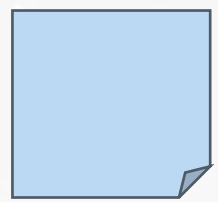


# TRACER: Brief Introduction

---

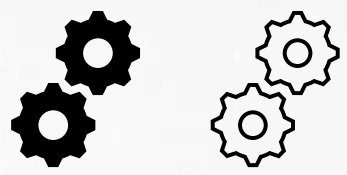
# Genesis of TRACER

Driven thru Configurable  
YAML



Track Multiple AI Projects  
across Releases

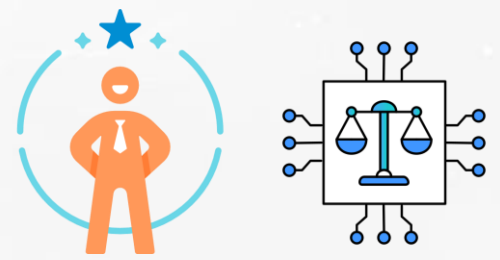
Configure Multiple Services



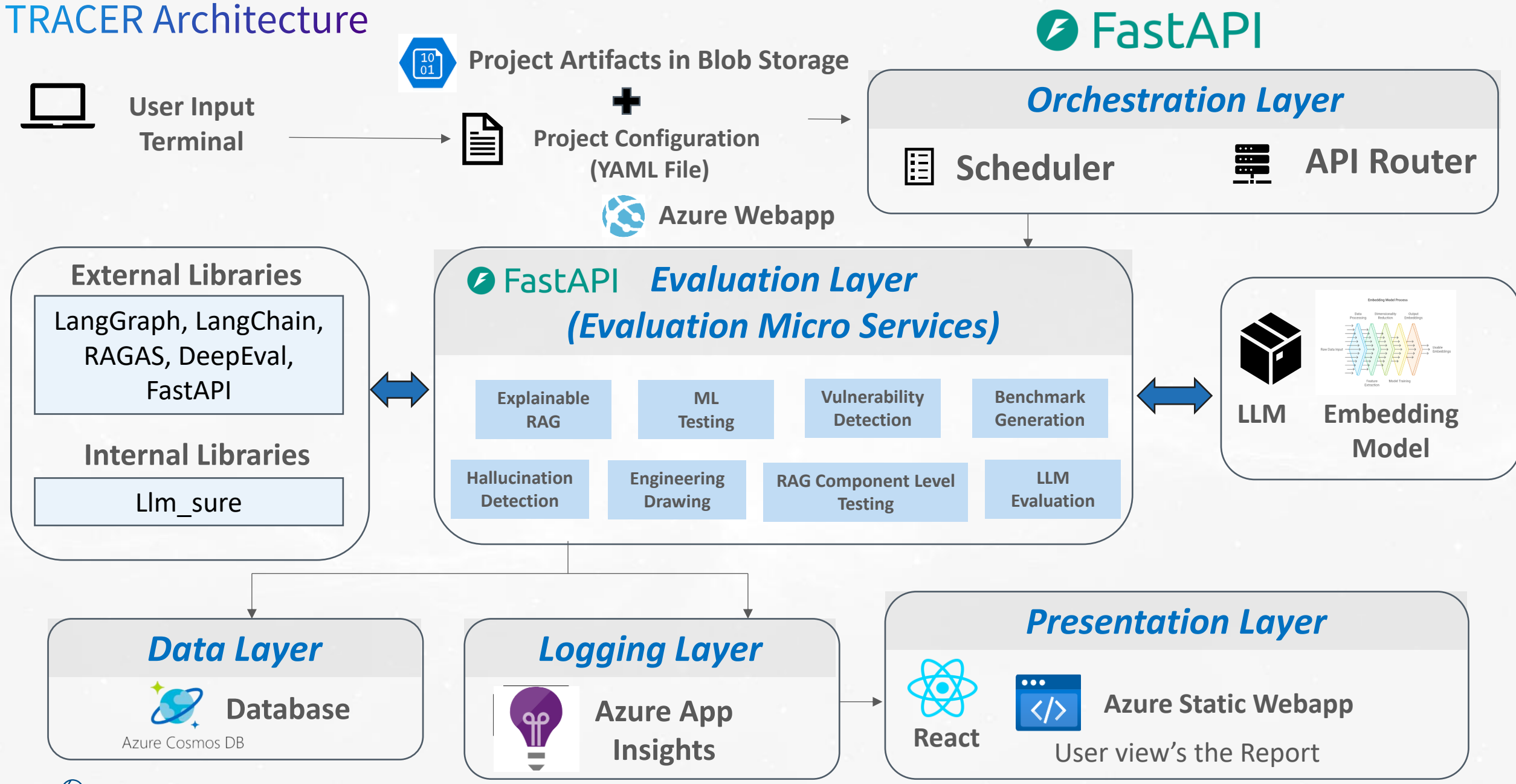
Broad Array of Evaluation  
Metrics



Responsible AI



# TRACER Architecture



# Our AI Assurance Offering

## 01 – Advisory & Consulting

- Holistic AI Assurance strategy for ensuring customer adoption and satisfaction

## 02 – Agent and App Validation

- Validating that the AI systems function in a robust, secure and safe way. They maintain Fairness at the highest levels and are transparent and explainable.

## 03 – Data and Model Validation

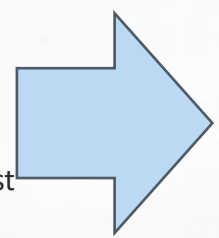
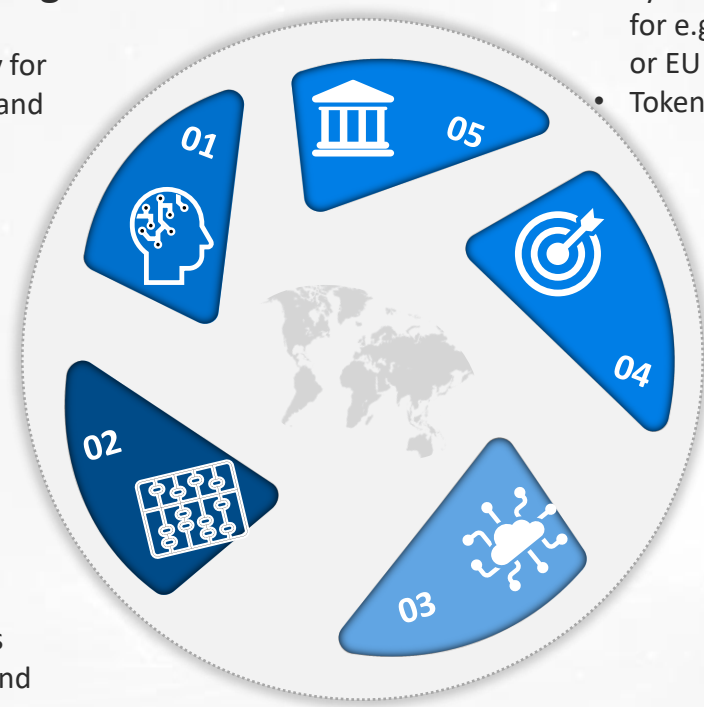
- Data Quality Check and Model validation for measurable quality, performance parameters and scalability

## 05 – Regulatory Compliance, Governance & Tokenomics

- Validating Compliance of AI Systems against specific Regulation for e.g. State/Country specific laws or EU AI Act etc
- Tokenomics

## 04 – Security

- Validating Artificial Intelligence systems, models, and data against potential threats, vulnerabilities, and malicious attacks maintaining CIA (Confidentiality, Integrity and Availability)



- Jump start on the AI Assurance Journey through TRACER
- TRACER – Trustworthy | Robustness | Accuracy | Compliance | Evaluation | and Reporting
- An end-to-end automated framework for "BLACK-BOX" testing of both Core and Generative AI

# TRACER – Trustworthy| Robustness| Accuracy| Compliance| Evaluation| and Reporting

Patent Filed in IN



LTIMindtree  
TRACER  
Framework

It provides an end-to-end automated framework for "BLACK-BOX" testing of both Core and Generative AI Applications, aimed at assessing Business Outcomes while encouraging Responsible AI practices

## APPROACH

To set up multiple Projects across various releases

To create and manage ground truth versions

To conduct automated testing using a wide range of innovative pre-built  
**EVALUATION FUNCTIONS**

To deliver explainable Reporting and Visualization

To provide Recommendations for identified issues

## EVALUATION FUNCTIONS

### Hallucination Detector

- Flexible & standardized framework to evaluate performance of chat systems and RAG based apps
- It detects hallucination probability of the system

### Explainable RAG

- This can evaluate any prompt & completion for RAG based apps in an Explainable manner
- To drive Adoption of RAG-based applications

### LLM Evaluation

- Input Prompt & Response Evaluation using both Qualitative & Quantitative Metrics are considered
- Evaluation in both Absence and Presence of Ground Truth

### LLM Vulnerability Detection

- Building Red teaming dataset for various vulnerabilities (e.g. toxicity, biasness etc.) to test readiness of the LLM apps
- Independent component, adaptable & provides recommendations

### Benchmark Dataset Generation

- Automatically generates comprehensive set of question-and-answer pairs from input documents towards ensuring the end-to-end evaluation of RAG-based applications.

### RAG-CLEF

- It's component level RAG evaluation framework
- Enables Targeted Optimization systems by identifying Strengths & Weaknesses at the Component Level

### ML Testing

- Agent driven testing of ML model life cycle.
- Identification and recommendations to correct issues of the ML pipeline.

### Engg Design Critique

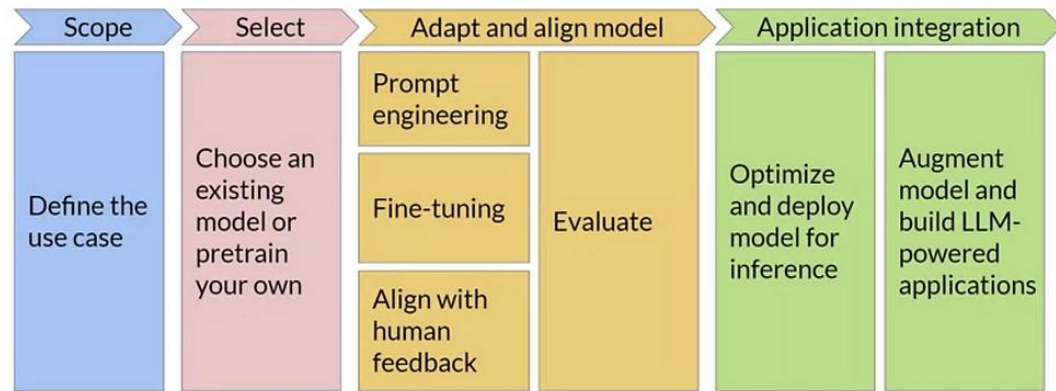
- A framework for analyzing & evaluating different types of Engineering Drawing (like part drawing, site plans etc.) .

LIVE & AVAILABLE for POC

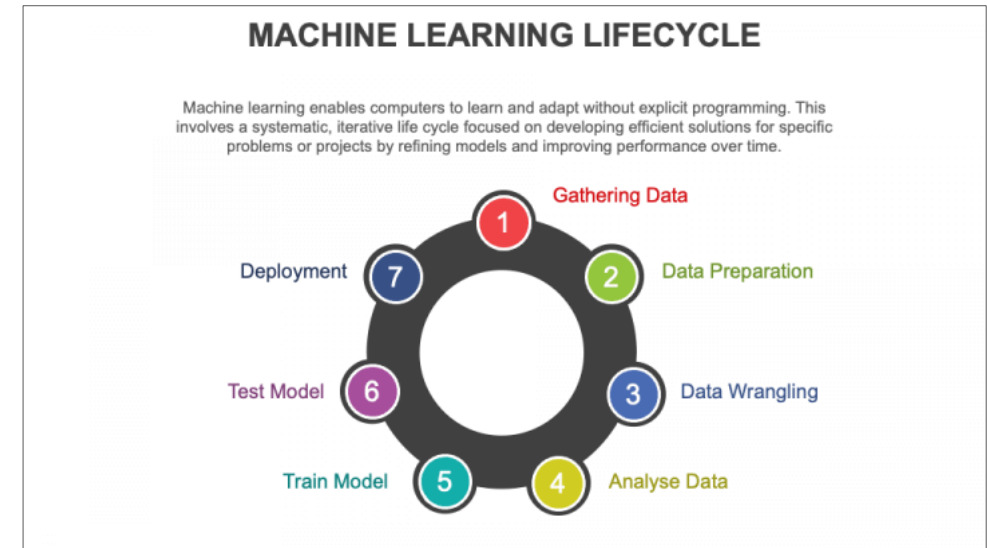
# Mapping between Responsible AI & TRACER

TRACER Feature	Responsible AI Dimension
LLM Evaluation, Vulnerability and Red Teaming	Ethical
Hallucination Detector, Explainable Evaluation of RAG based Application, LLM Evaluation, Vulnerability and Red Teaming, Benchmark Dataset Generation, RAG Component Level Testing	Explainable, Ethical, Efficient, Comprehensive
Hallucination Detector	Explainable
LLM Evaluation, Compliance Checker	Ethical, Efficient
LLM Evaluation	Efficient

# Applying TRACER during Iterative Development or Production



Gen AI Life Cycle: Apply TRACER during Development for Evaluation and Production for further Fine Tuning

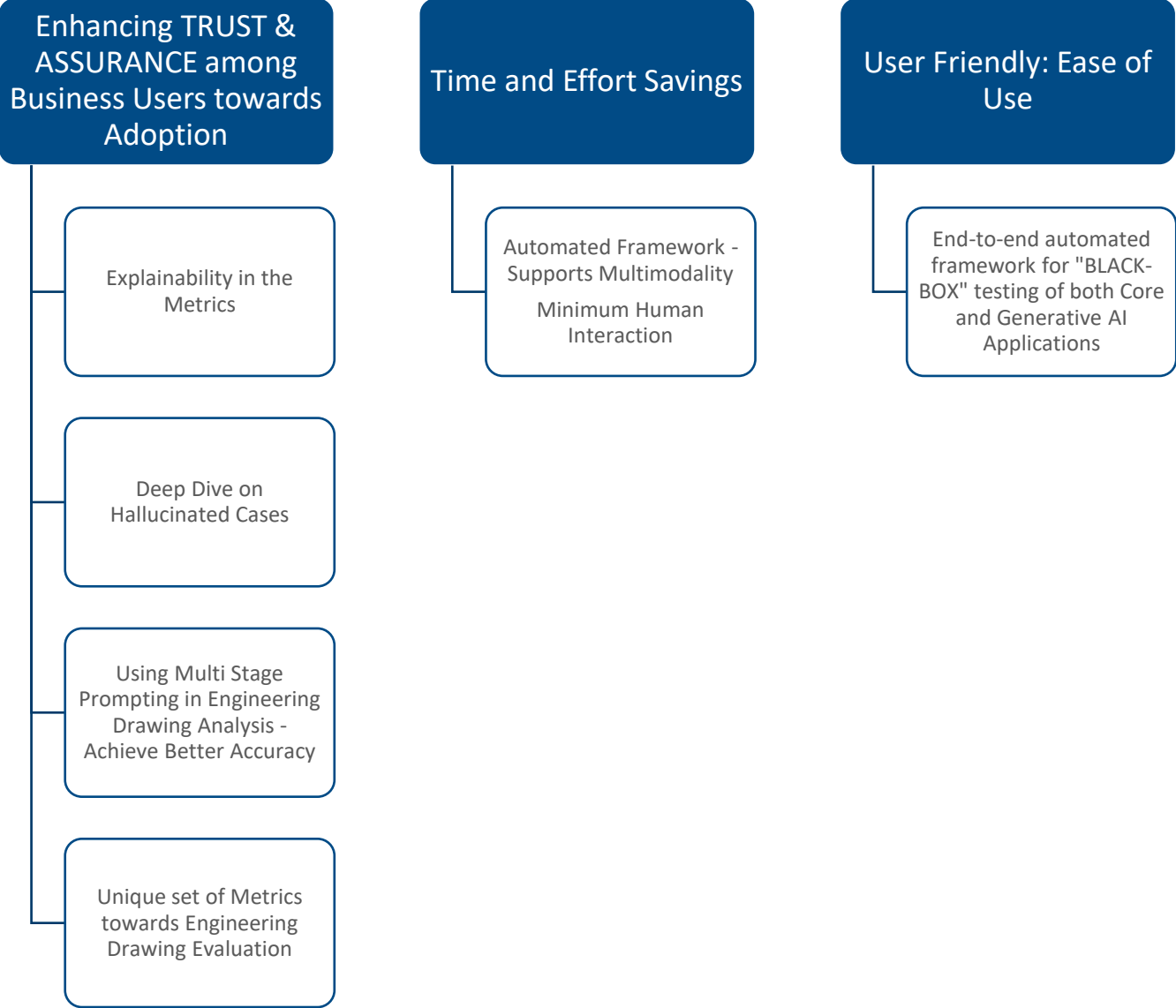


Recommendations about:

- Requirement Analysis
- Data Analysis
- Feature Engineered Data Analysis
- Train Test Distribution Mismatch
- Model Analysis
- Hyper-parameter Tuning



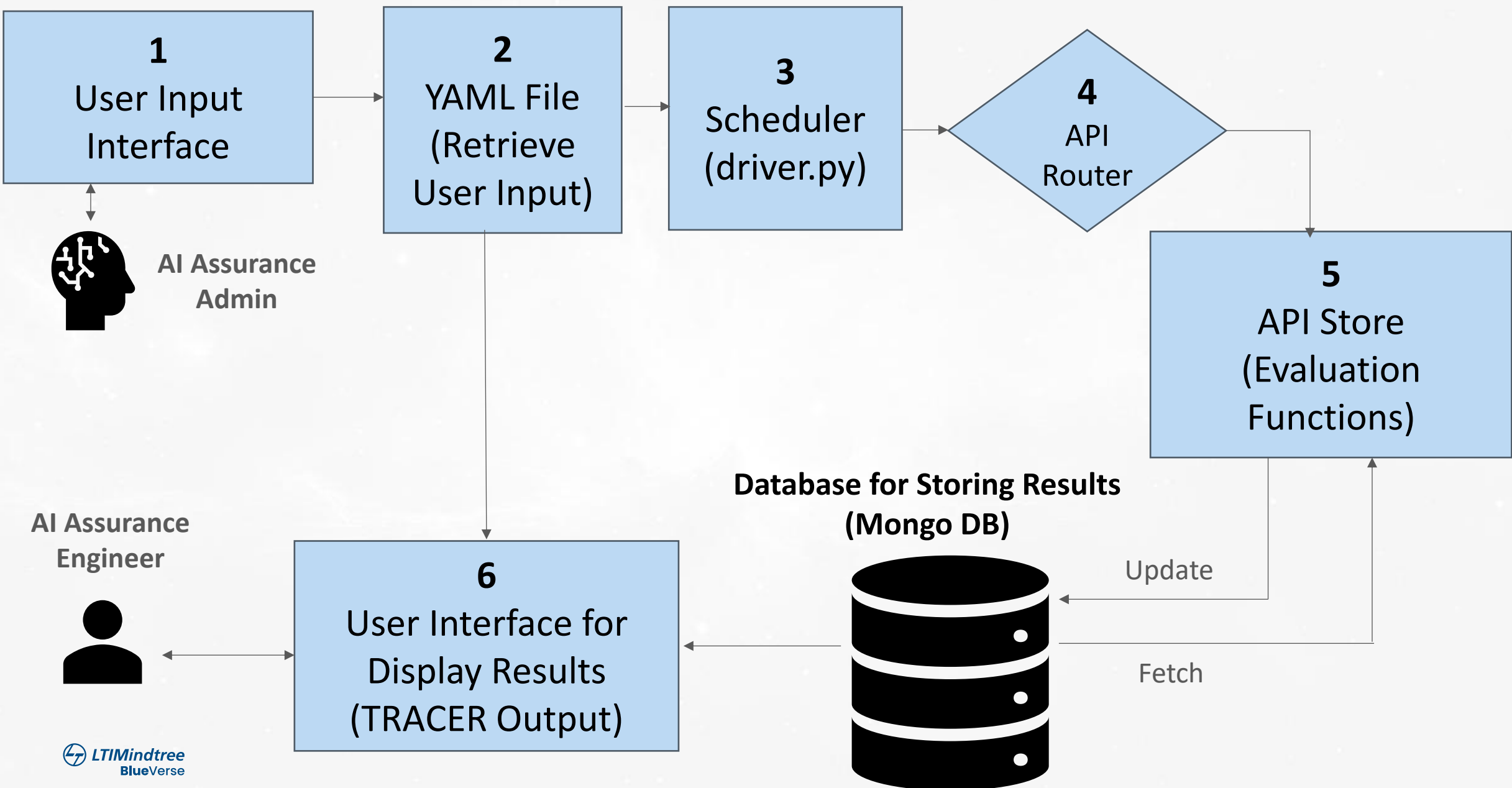
# Key Takeaways from TRACER



The advantages of the proposed solution can be measured from three different dimensions which are presented here. The figure depicts various factors behind the improvement for each dimension.



# TRACER Process Flow Diagram



# TRACER – User view

Select the Project

Project-1

Select the Release Version

1.0

Select the Accelerator

Hallucination Detection

Display Result

☐ Display Detail Result

☐ Evaluation Metric

TRACER

Trustworthy Robustness Accuracy Compliance Evaluation and Reporting

Evaluation Summary

	Prompt	Ground Truth
0	What are the disadvantages of LSTMs?	LSTMs still struggle with the problem of sequential processing, w
1	what is the similarity in terms of architecture when transformer mod	The similarity in terms of architecture when transformer models e
2	what is mlp?	The multilayer perceptron (MLP) is a classic form of neural netwo
3	how is lstm architecture is applied in speech processing?	The provided context does not explicitly detail the application of l
4	what is transformer in electrical?	The given context does not provide information about transforme
5	What is Transformer?	Transformers are a type of deep neural network (DNN) that addre

Select the Project

Project-1

Select the Release Version

1.0

Select the Accelerator

Hallucination Detection

Display Result

☒ Display Detail Result

☒ Evaluation Metric

Evaluation progress: 10/10

Common Metrics

BLEU Score(%)

46.82

ROUGUE-2 Score(%)

53.06

ROUGUE-L Score(%)

63.91

Answer Relevancy(%)

56.14

Answer Correctness(%)

79.84

Answer Similarity(%)

97.54

Response Relevancy(%)

Faithfulness(%)

Noise Sensitivity(%)

Select the Project

Project-1

Select the Release Version

1.0

Select the Accelerator

Hallucination Detection

Display Result

☒ Display Detail Result

☒ Evaluation Metric

Hallucination Rate: 40.0%

Claims Similarity(%)

Percentage of claims in Response inferred from Ground Truth

17.5

Truthful Claims(%)

Percentage of claims in Response inferred from Retrieved Context

38.33

Relevant Claims(%)

Percentage of claims in Response inferred from Question

26.67

Claims Gap(%)

Percentage of claims in Response NOT inferred from Ground Truth

82.5

False Claims(%)

Percentage of claims in Response NOT inferred from Retrieved Context

61.67

Irrelevant Claims(%)

Percentage of claims in Response NOT inferred from Question

73.33

# TRACER – User view

The screenshot shows the TRACER web application interface. On the left, there are dropdown menus for 'Select the Project' (Project-1), 'Select the Release Version' (1.0), and 'Select the Accelerator' (Hallucination Detection). Below these is a 'Display Result' button and checkboxes for 'Display Detail Result' and 'Evaluation Metric'. The main area displays the title 'TRACER' and the subtitle 'Trustworthy Robustness Accuracy Compliance Evaluation and Reporting'. Below this is an 'Evaluation Summary' table with two columns: 'Prompt' and 'Ground Truth'.

	Prompt	Ground Truth
0	What are the disadvantages of LSTMs?	LSTMs still struggle with the problem of sequential processing, w
1	what is the similarity in terms of architecture when transformer mod	The similarity in terms of architecture when transformer models e
2	what is mlp?	The multilayer perceptron (MLP) is a classic form of neural netwo
3	how is lstm architecture is applied in speech processing?	The provided context does not explicitly detail the application of l
4	what is transformer in electrical?	The given context does not provide information about transforme
5	What is Transformer?	Transformers are a type of deep neural network (DNN) that addre

The screenshot shows the TRACER web application interface with the 'Common Metrics' section. It displays an 'Evaluation progress: 10/10' bar. Below this, there are six green boxes showing various metrics:

Metric	Value
BLEU Score(%)	46.82
ROUGUE-2 Score(%)	53.06
ROUGUE-L Score(%)	63.91
Answer Relevancy(%)	56.14
Answer Correctness(%)	79.84
Answer Similarity(%)	97.54

The screenshot shows the TRACER web application interface with the 'Hallucination Rate: 40.0%' section. It displays six blue boxes showing various metrics:

Metric	Value
Claims Similarity(%)	17.5
Truthful Claims(%)	38.33
Relevant Claims(%)	26.67
Claims Gap(%)	82.5
False Claims(%)	61.67
Irrelevant Claims(%)	73.33

Getting to the  
***Future. Faster.***  
Together.

A vibrant nebula with blue, purple, and orange hues against a dark starry background. The nebula is located on the right side of the image, with a bright orange and yellow core. The text "Getting to the Future. Faster. Together." is overlaid on the left side of the image. The word "Future" is in a bold, italicized, light blue font. The word "Faster" is in a bold, italicized, light blue font. The word "Together" is in a light blue font. A dotted line extends from the end of the word "Together" to a solid blue circle.