

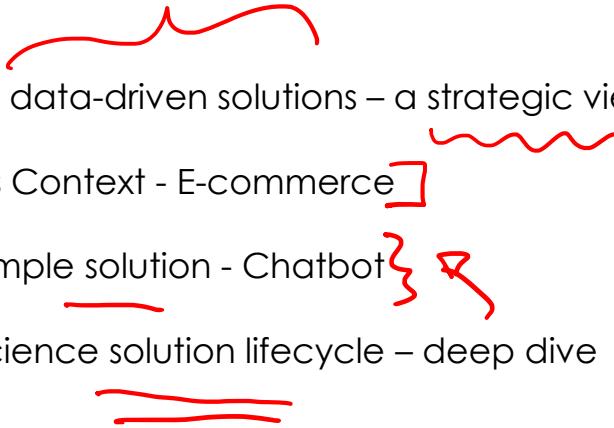
The Art and Science of Data

Navigating the Data Science Lifecycle

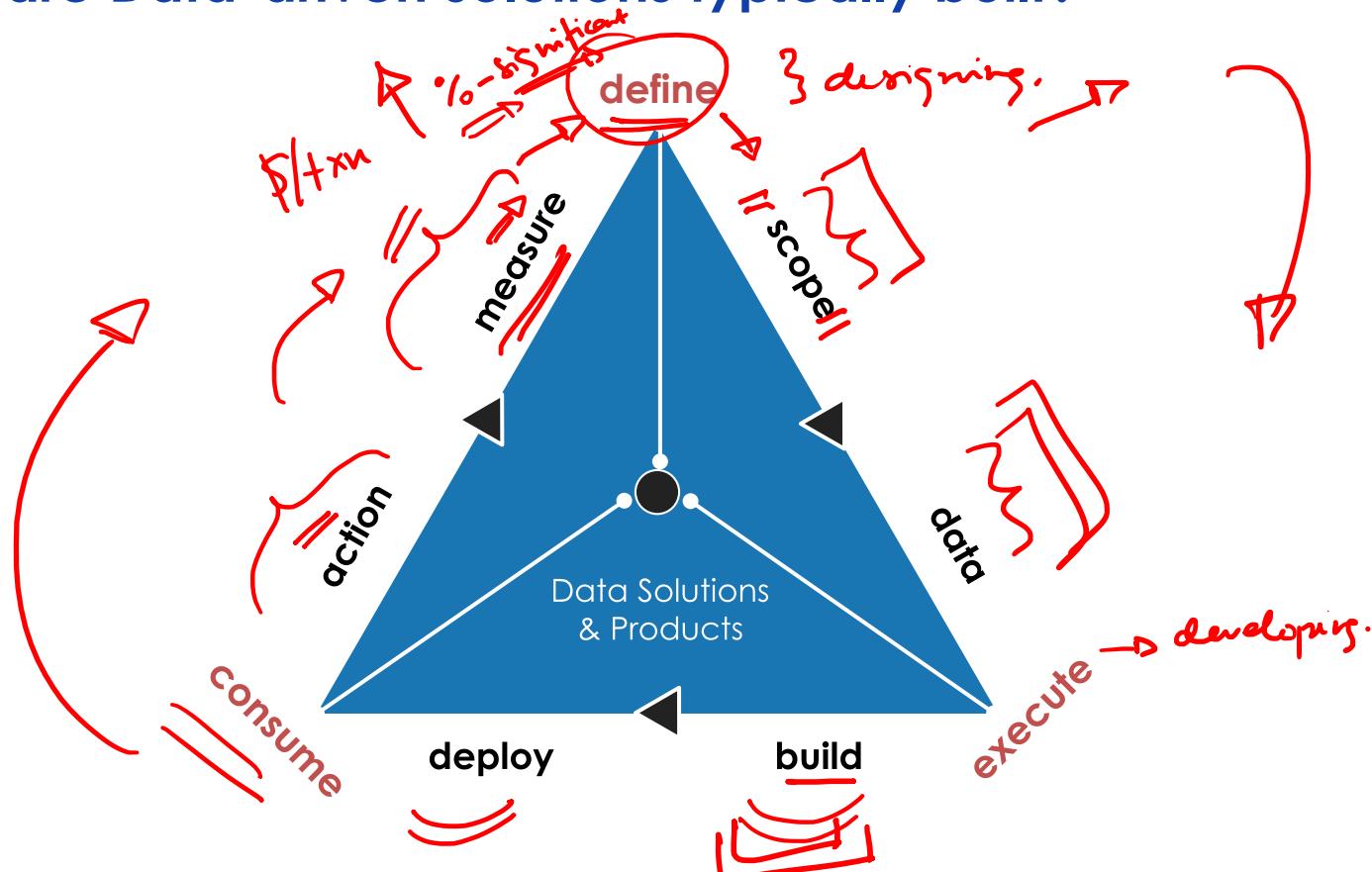
Learning Outcomes

- Understanding of the data science and machine learning lifecycle
- Understanding of the key considerations in navigating a data science lifecycle
- Preliminary insights into underlying architecture that powers a data science solution
- Comprehension of different job roles coming together to deliver these solutions

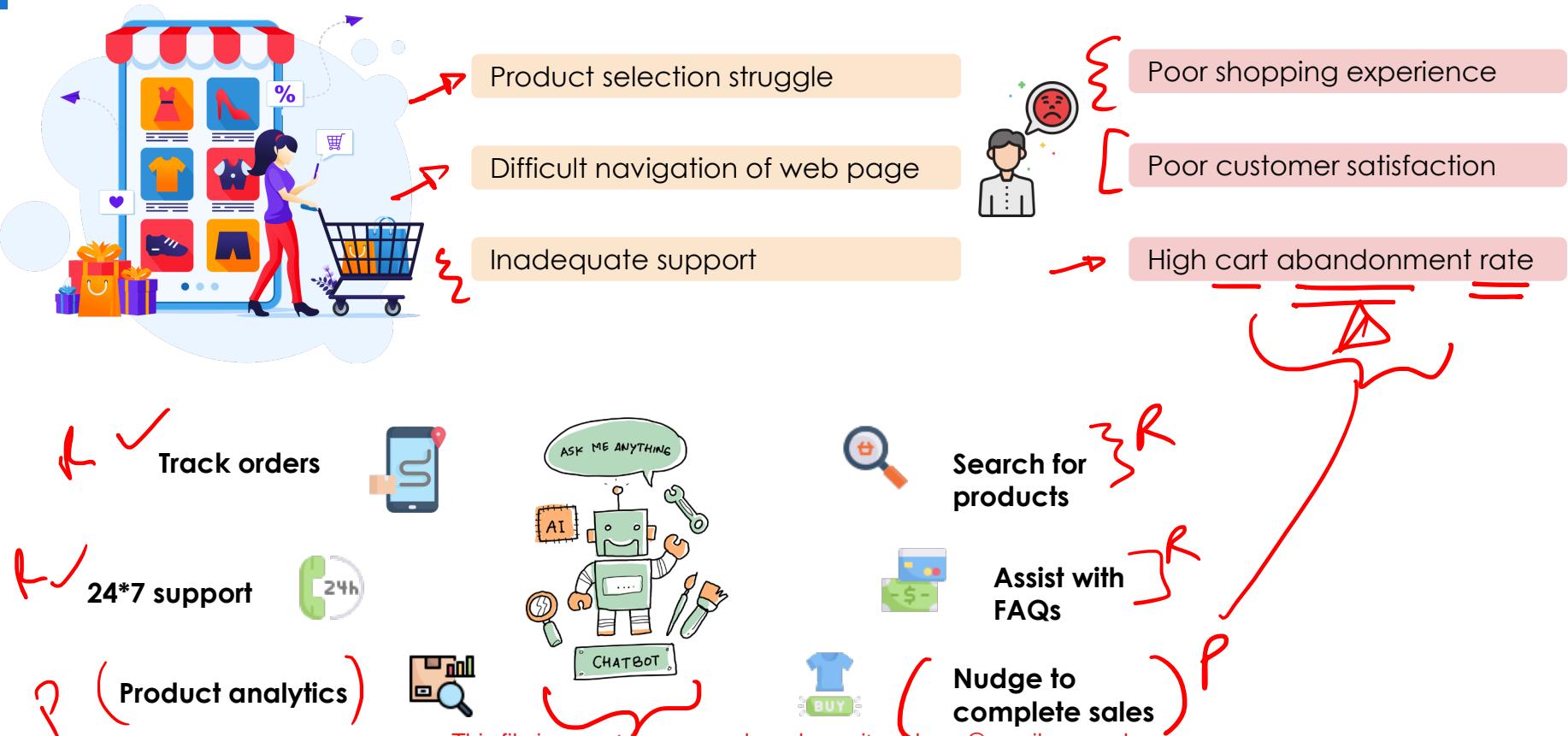
Agenda

- Building data-driven solutions – a strategic view
 - Business Context - E-commerce
 - An example solution - Chatbot
 - Data Science solution lifecycle – deep dive
- 

How are Data-driven solutions typically built?



Business Context - E-commerce

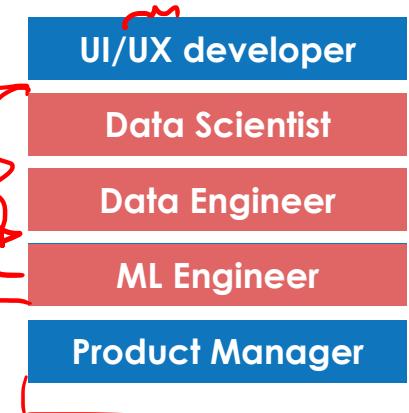


Strategy for building the ChatBot

“If you were owning this business, what teams would have to come together to build this solution?”

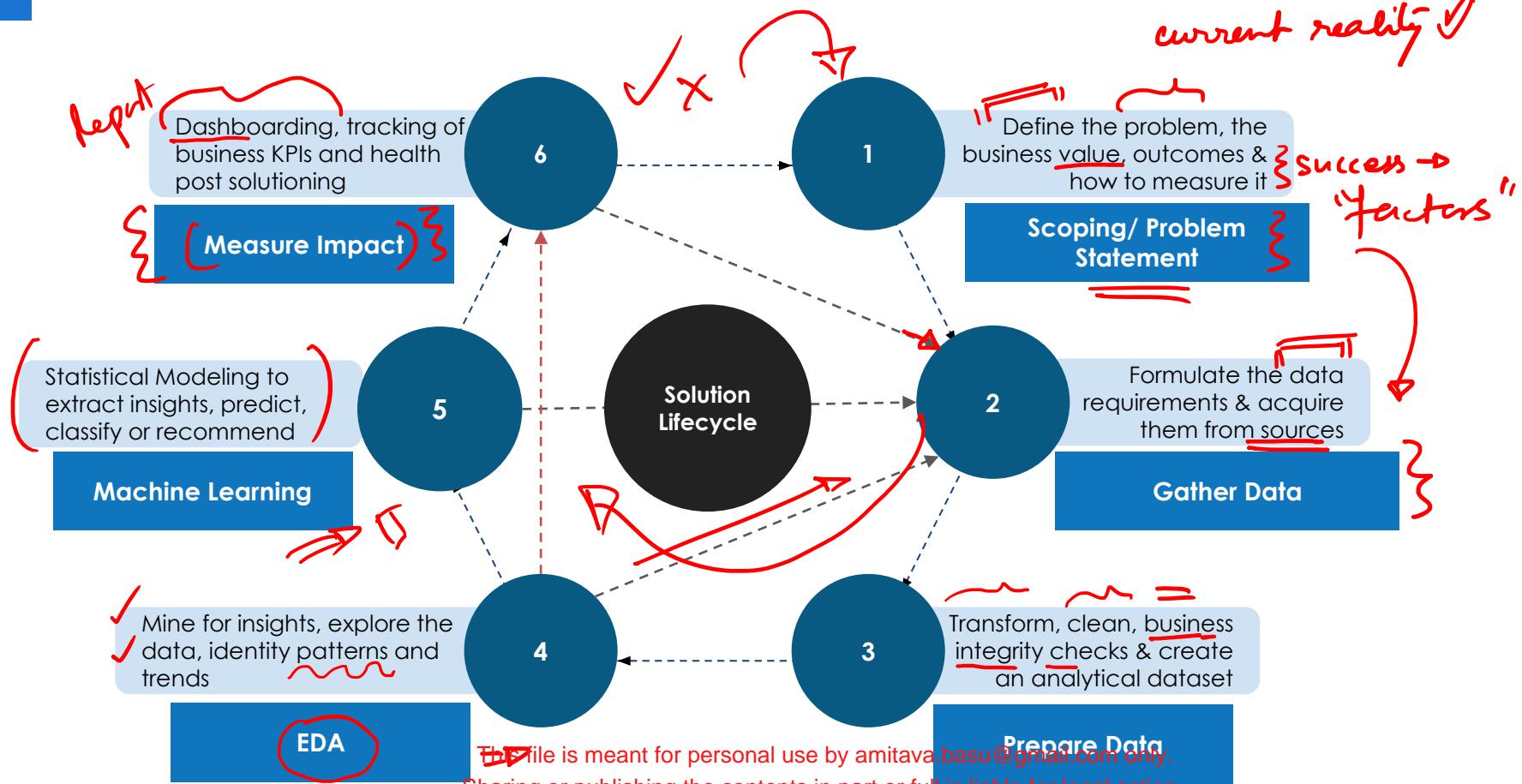
You would need someone to -

- Create the UI of the chatbot and integrate it into your website.
- Create a language based model to answer your customers' questions.
- Create/gather and manage the data that is needed to train your model.
- Deploy that model, monitor it and manage its performance.
- Manage and drive the development & consumption of this product.

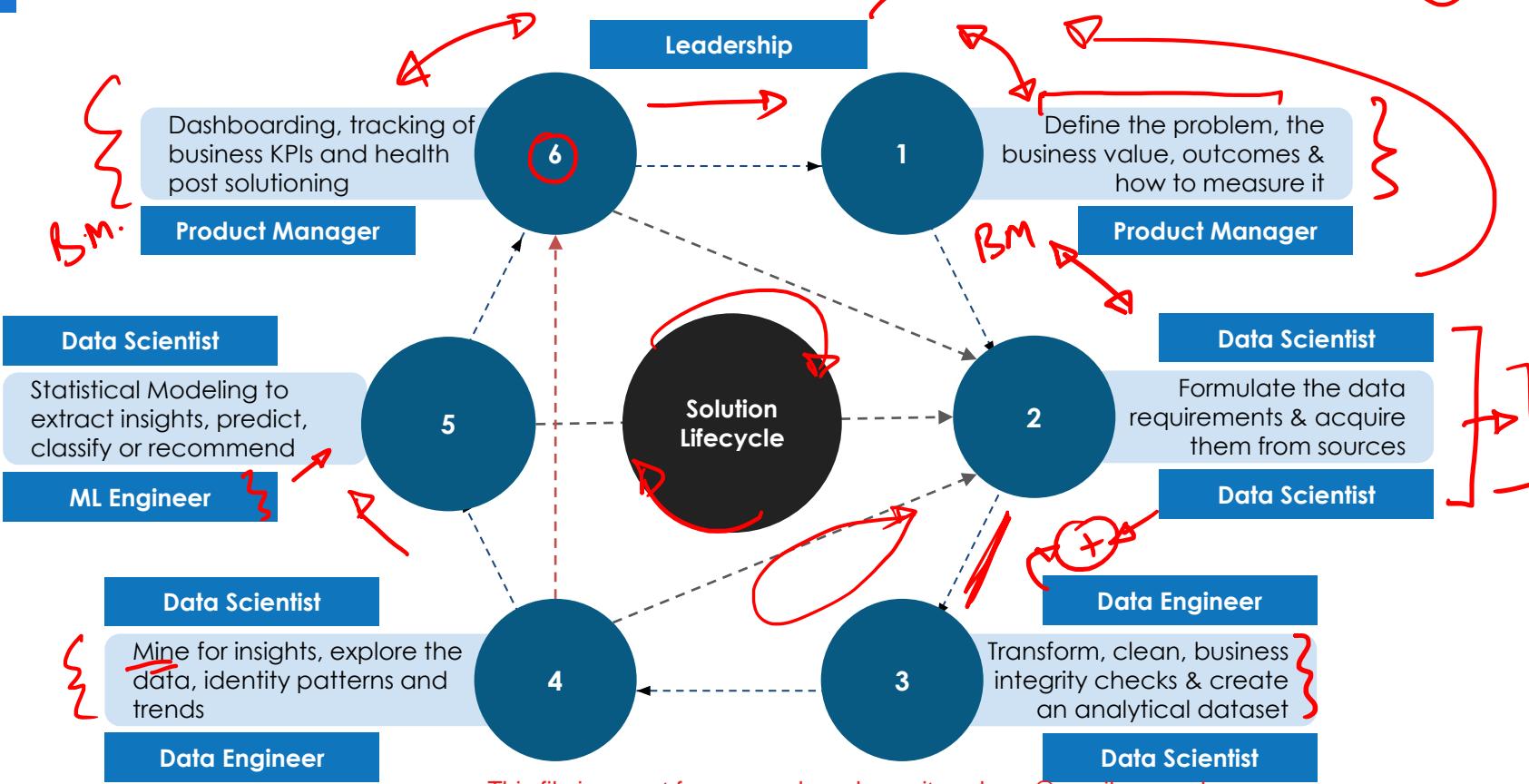


COLLABORATION
ChatBot is built!

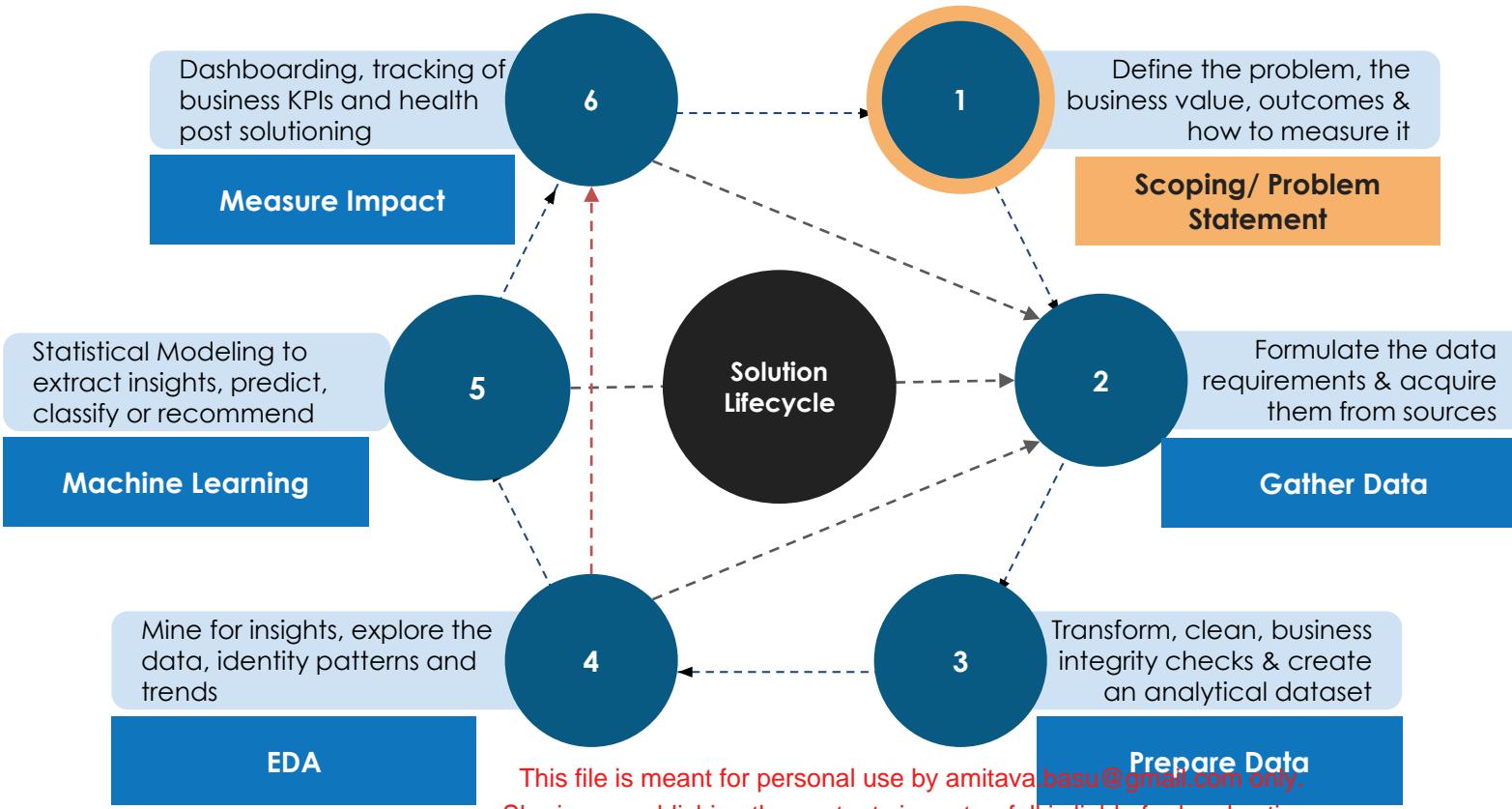
A Typical Solution Lifecycle



A Typical Solution Lifecycle

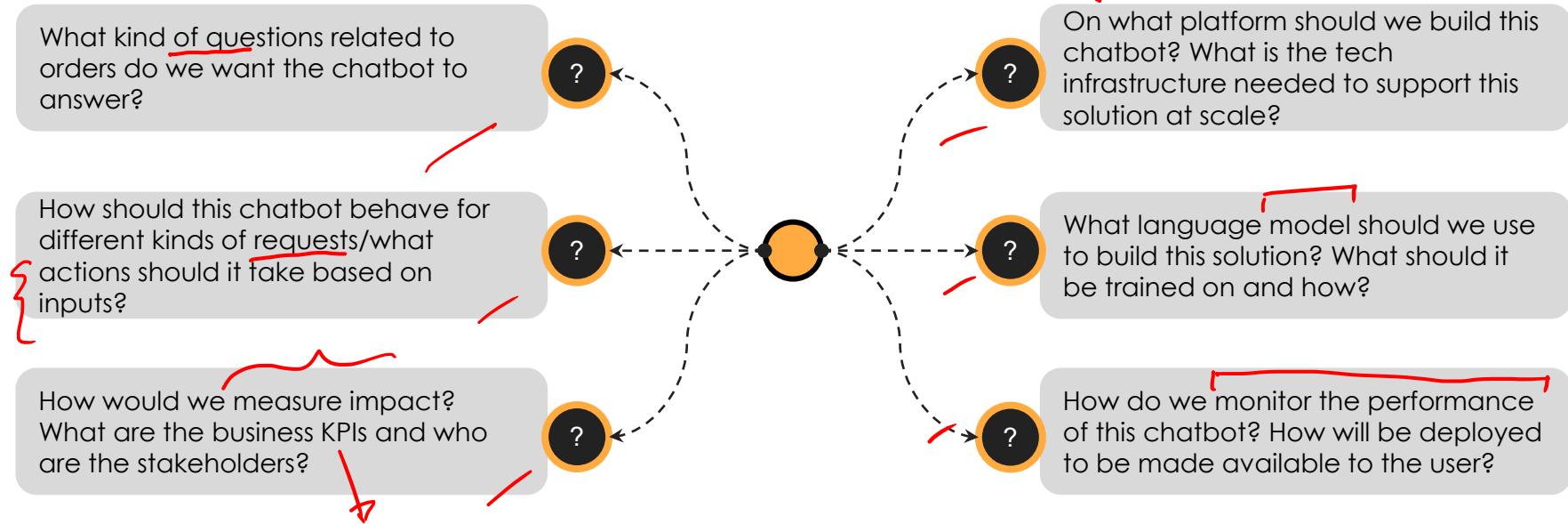


A Typical Solution Lifecycle

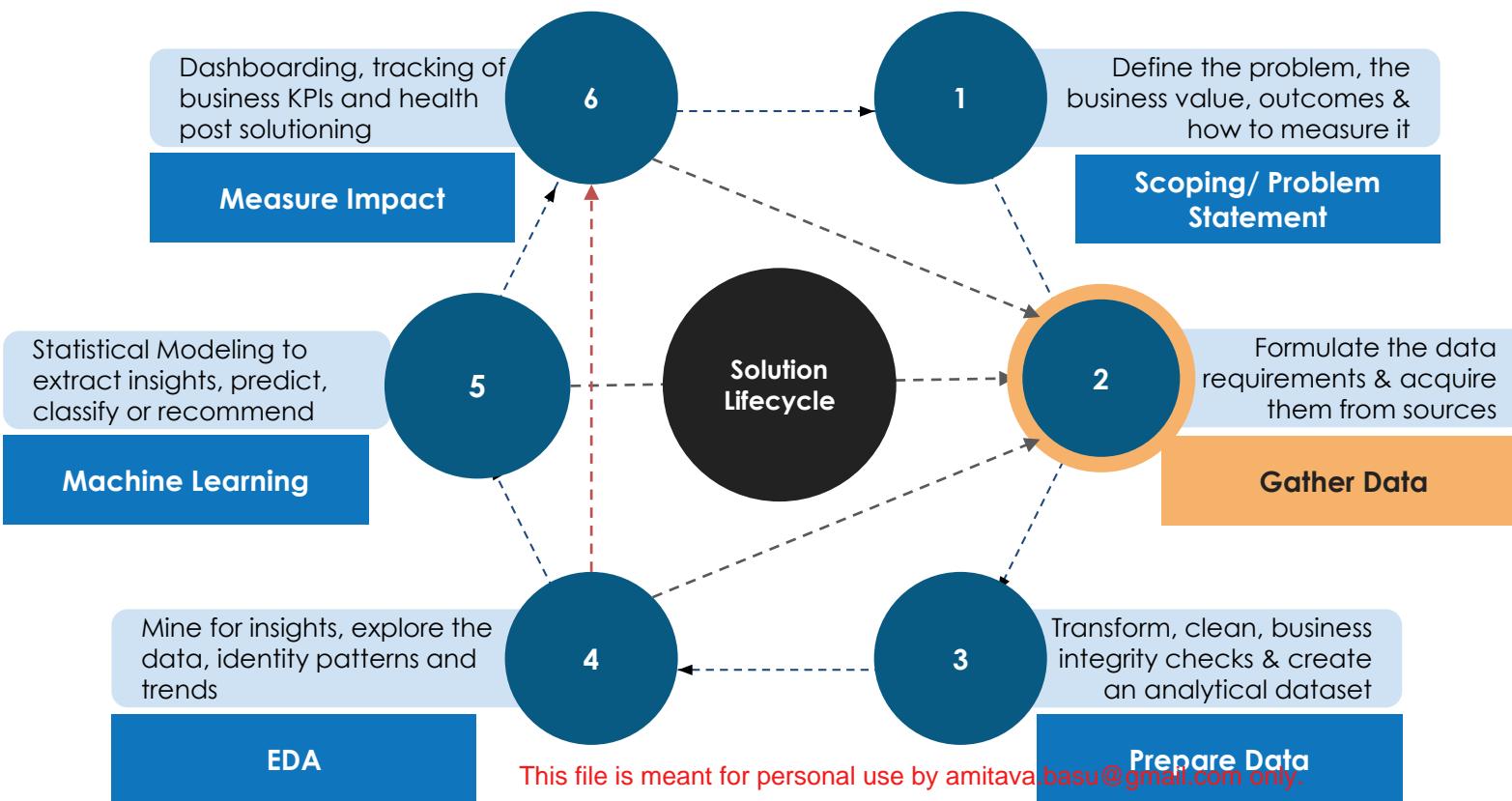


Scoping/Problem Definition

“What should this ChatBot be able to do/how would it work?”



A Typical Solution Lifecycle



Gather Data

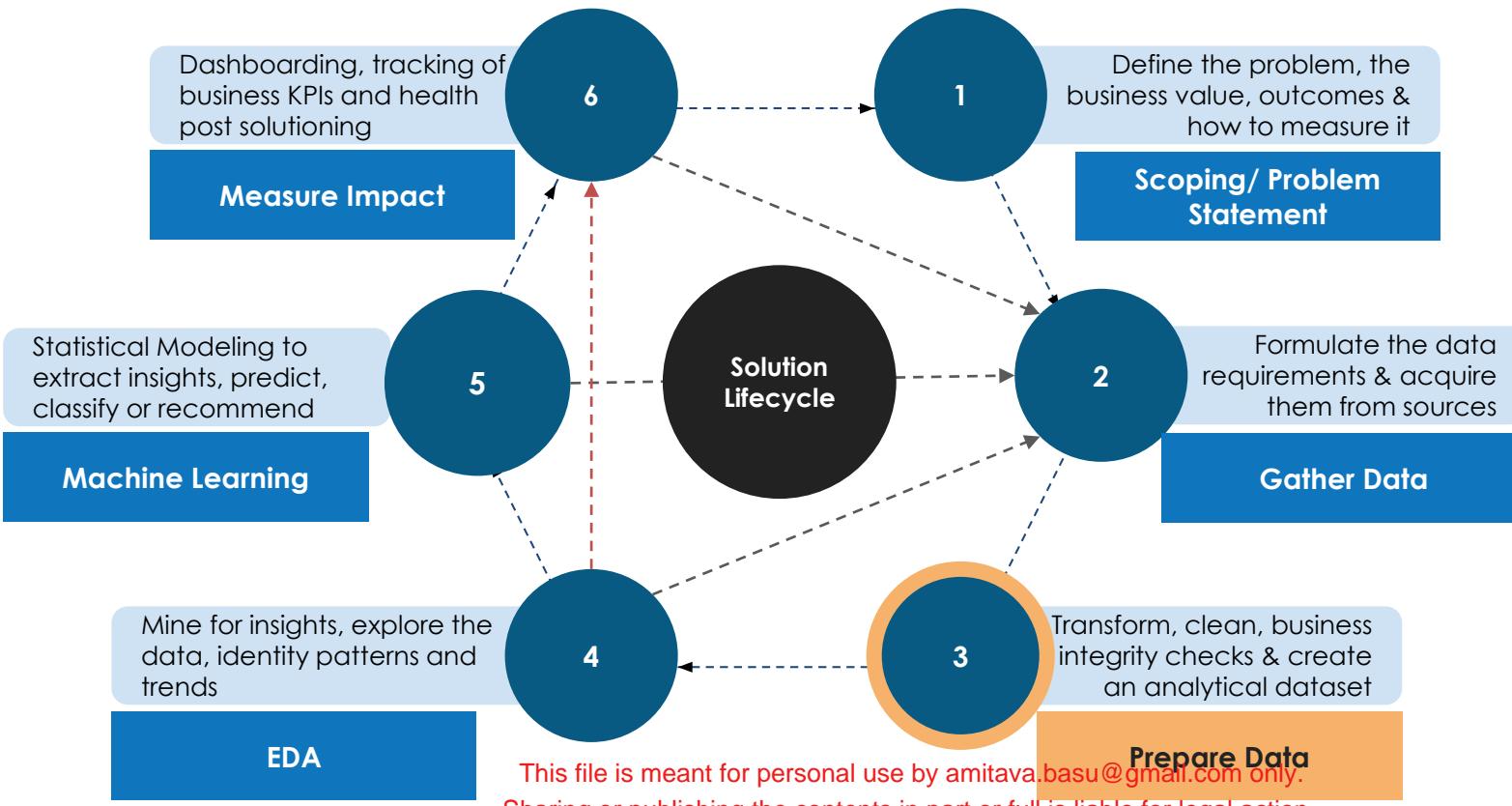
“What data do I need to build this chatbot?”



This file is meant for personal use by amitava.basu@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action

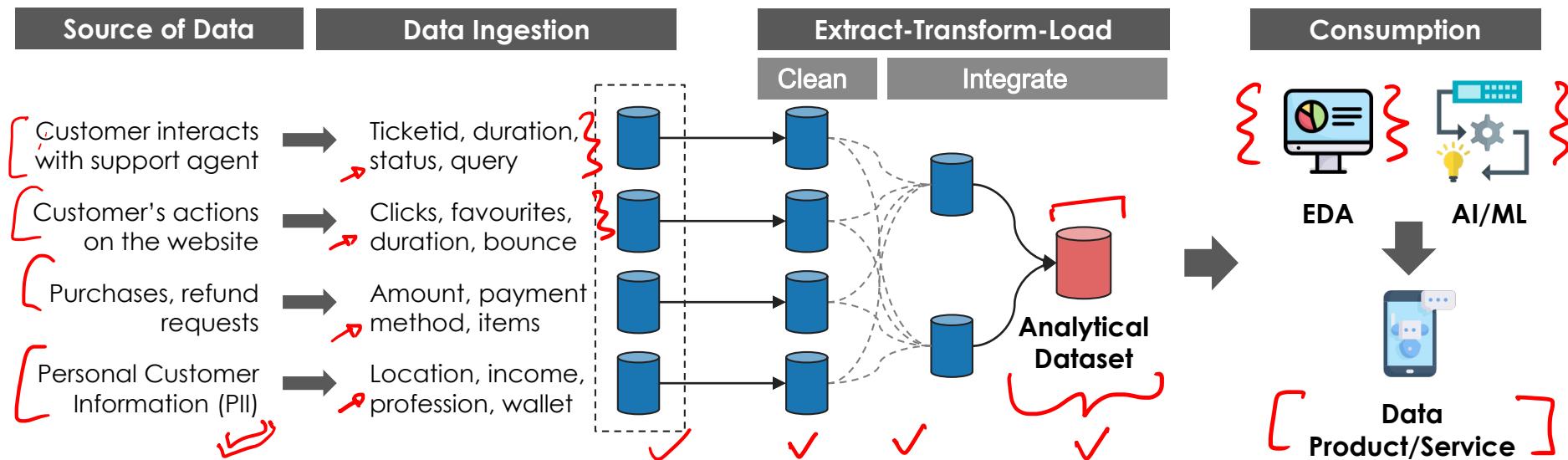
Proprietary Content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

A Typical Solution Lifecycle



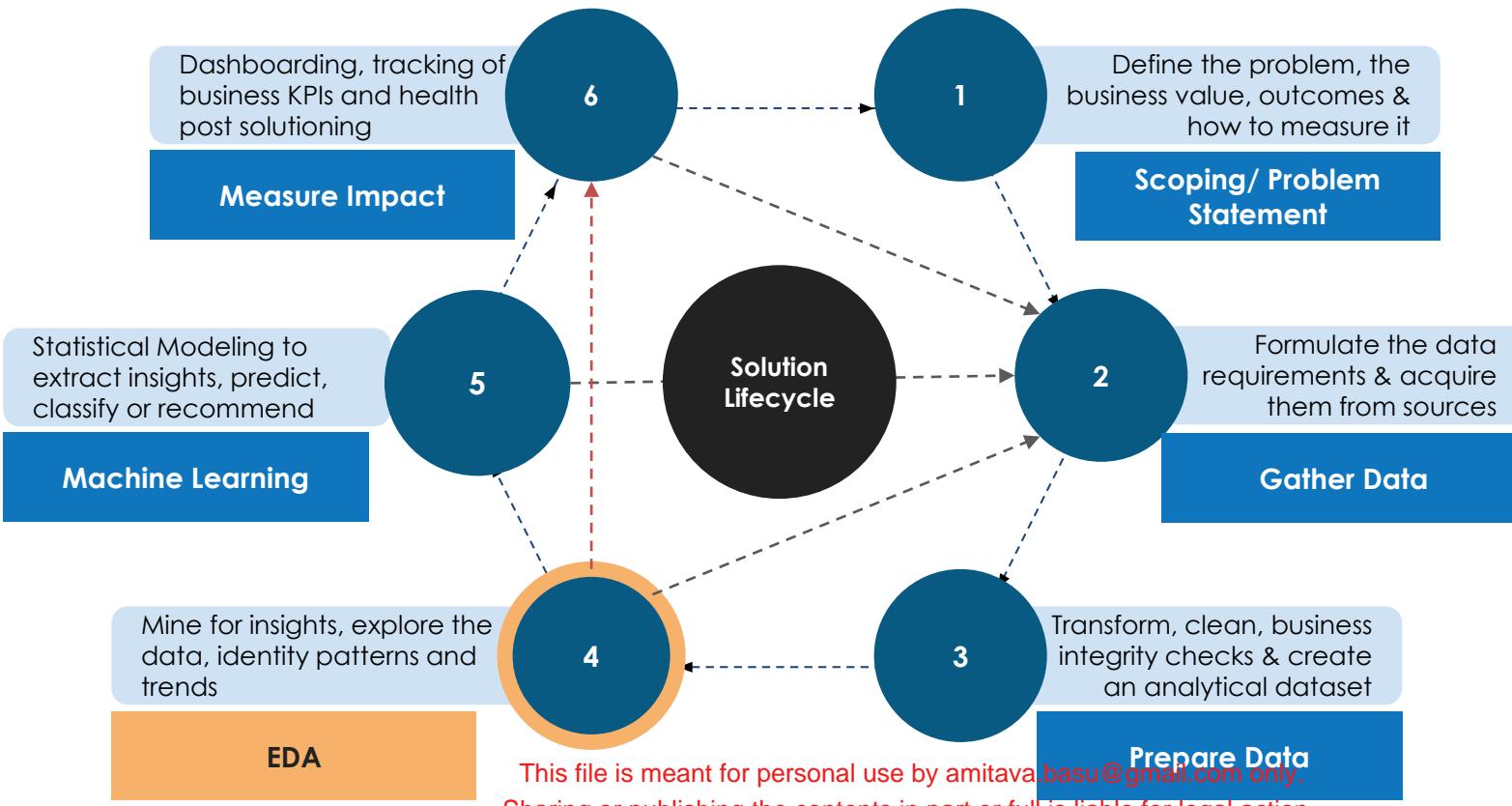
Prepare Data

“How do I create the data I need for my data solution/product?”



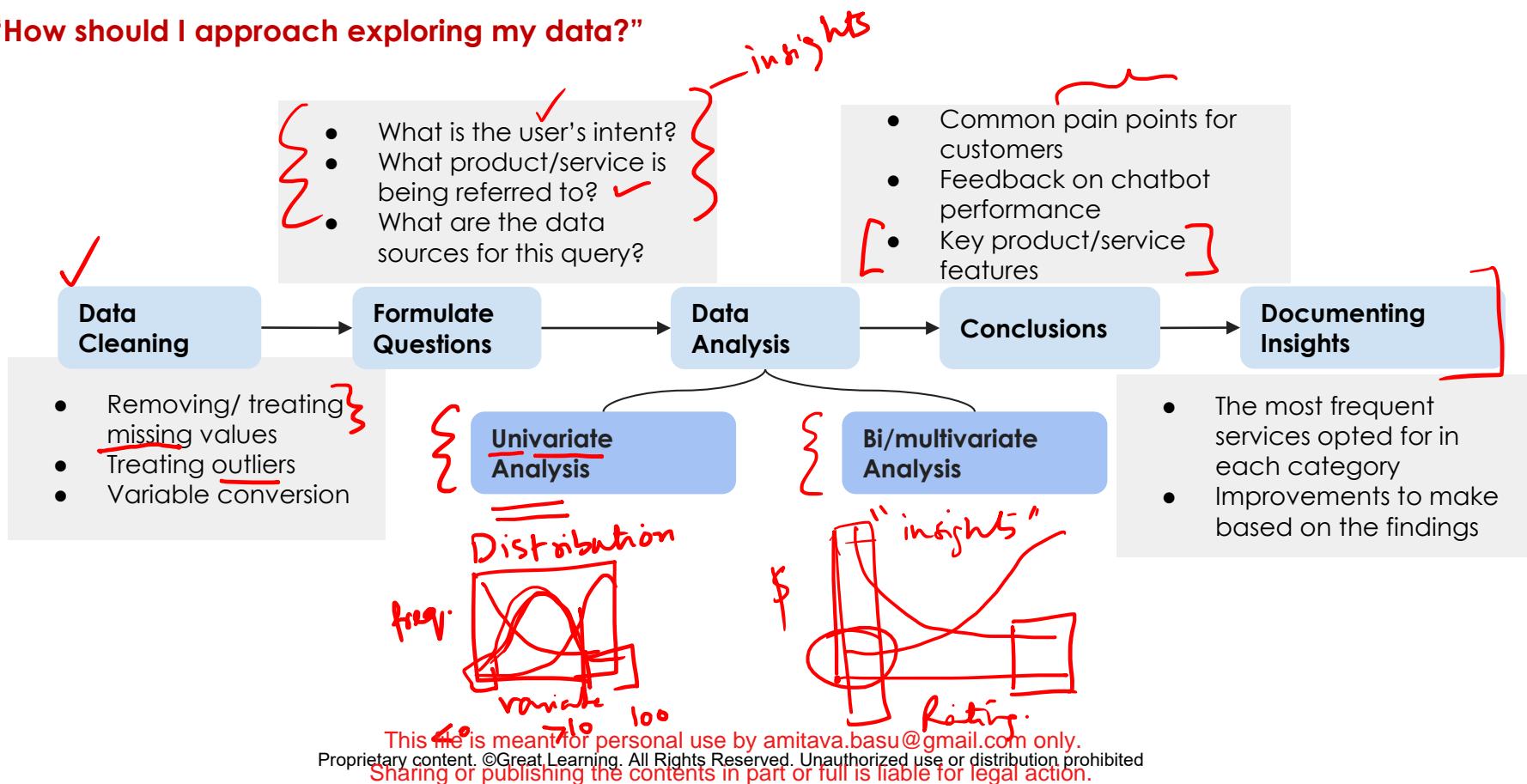
Various databases within a data warehouse like **MySQL**

A Typical Solution Lifecycle

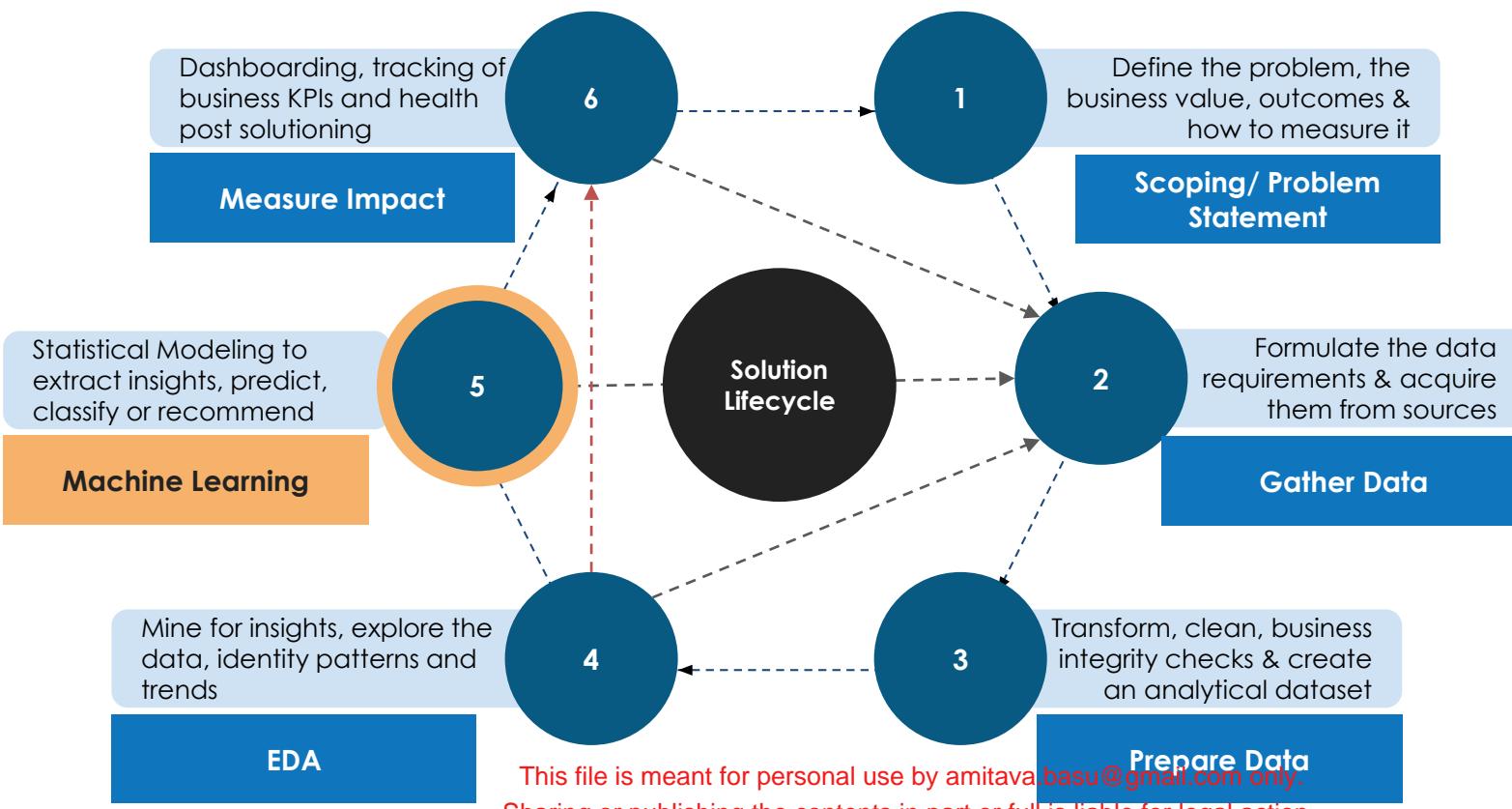


Exploratory Data Analysis

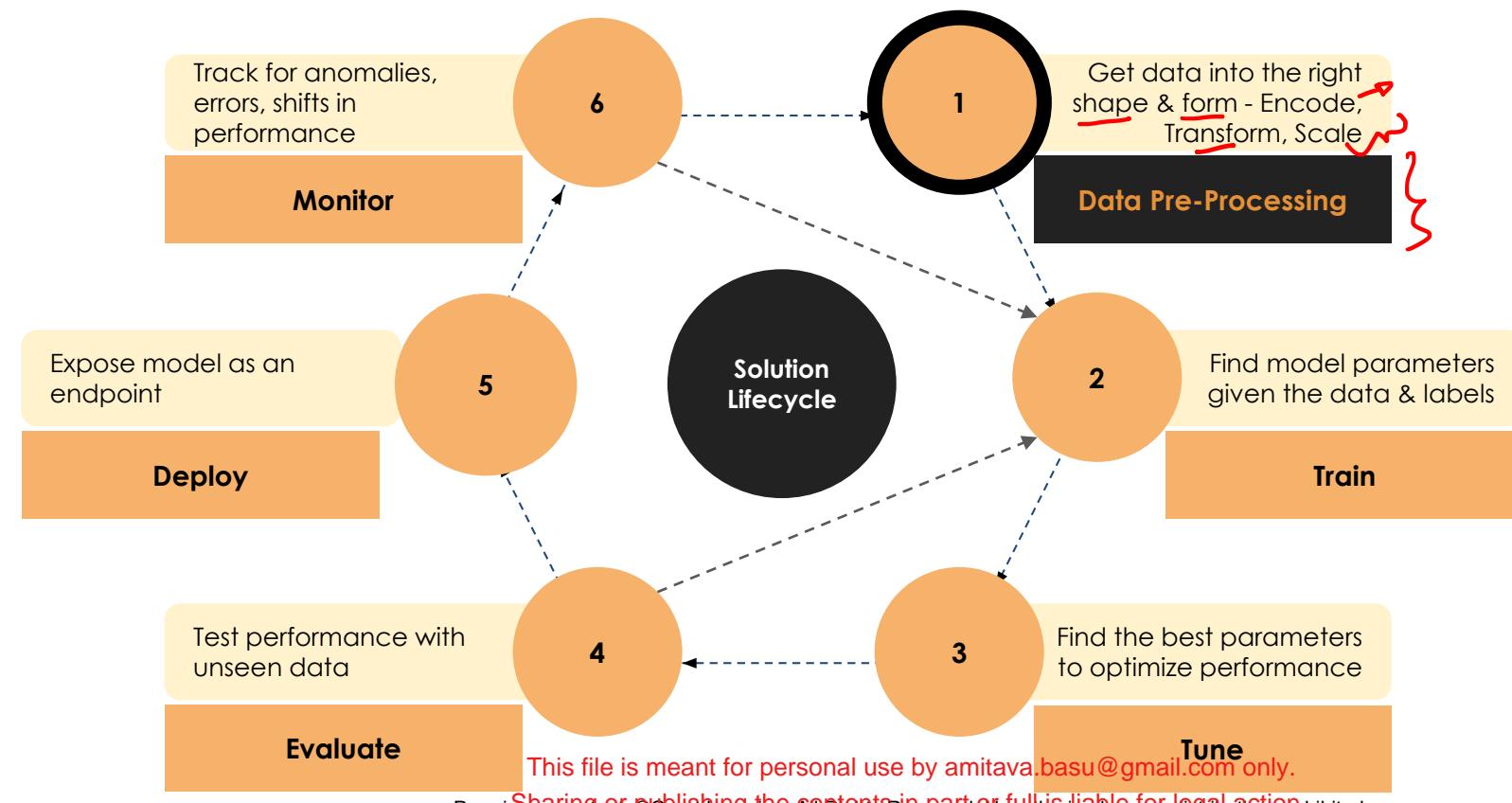
“How should I approach exploring my data?”



A Typical Solution Lifecycle



A Typical Solution Lifecycle



Data Pre-processing



How does the chatbot understand us?

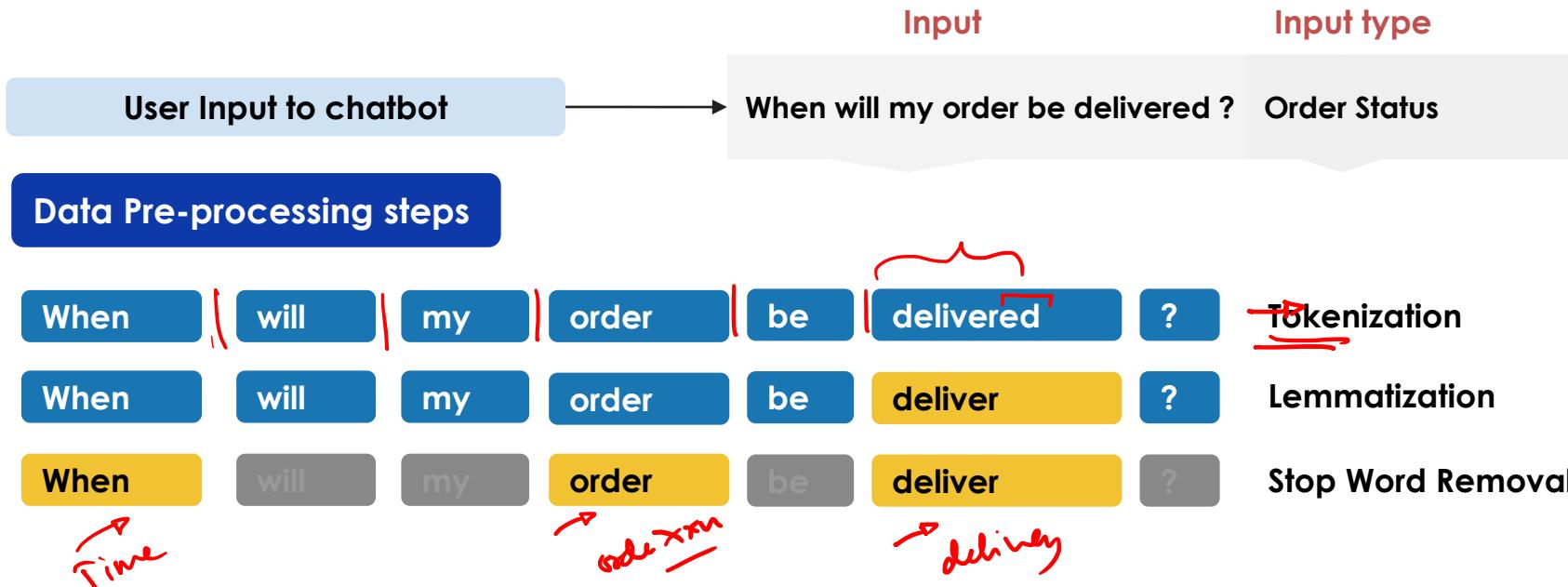
I want to buy a new HP Laptop	Make a purchase	Intent Recognition
I want to buy a <u>new</u> <u>HP</u> Laptop	Product - new Laptop	Entity Recognition
I want to buy a new <u>HP</u> Laptop	Brand - HP	Named Entity Recognition
I want to buy a new HP Laptop	Sentiment - Positive	Sentiment Analysis

This is how a machine understands - This is a customer, who seems to have an overall positive sentiment, and wants to purchase a new HP laptop.

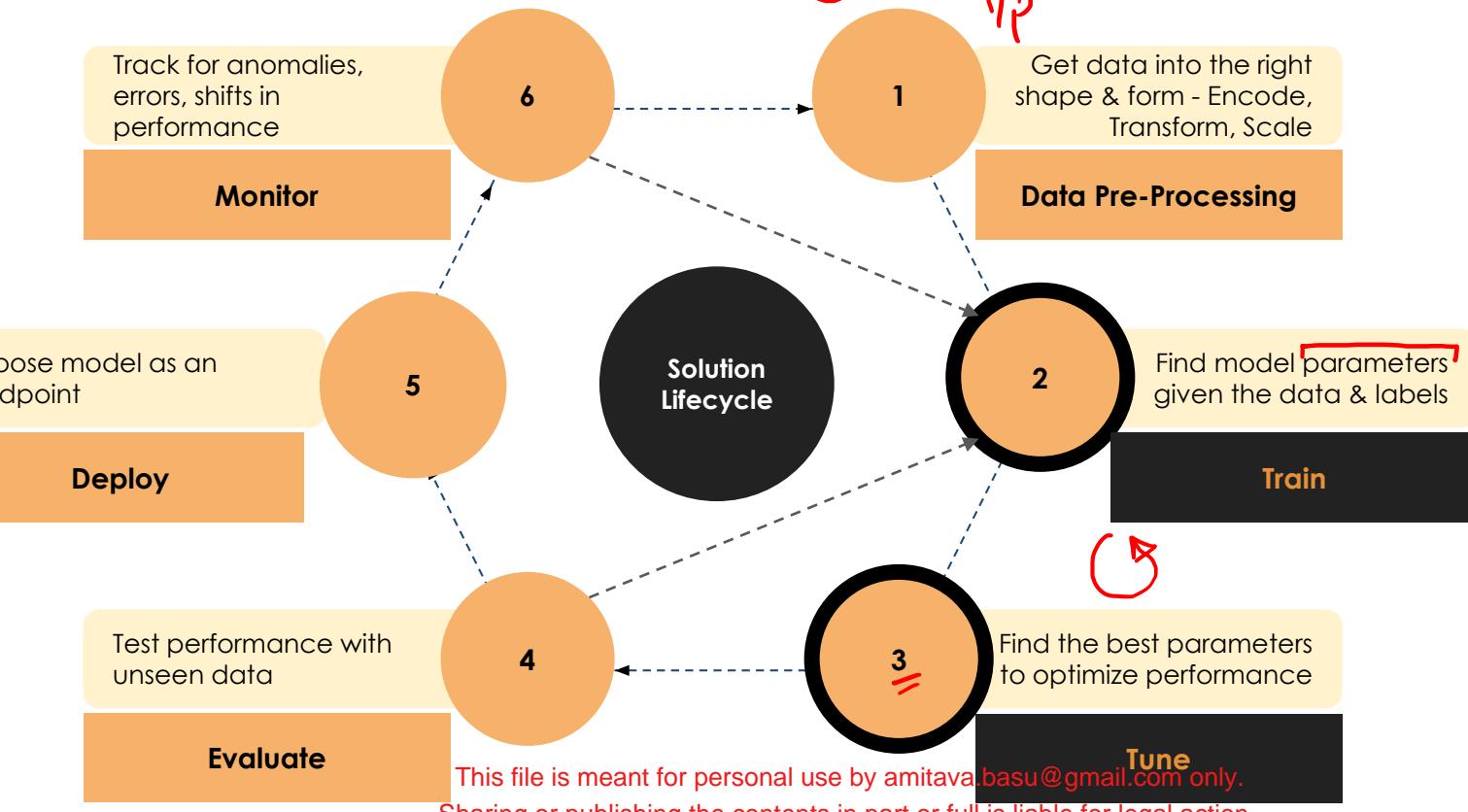
This file is meant for personal use by amitava.basu@gmail.com only.

Sharing or publishing the contents in part or full is liable for legal action.

Data Pre-processing



A Typical Solution Lifecycle

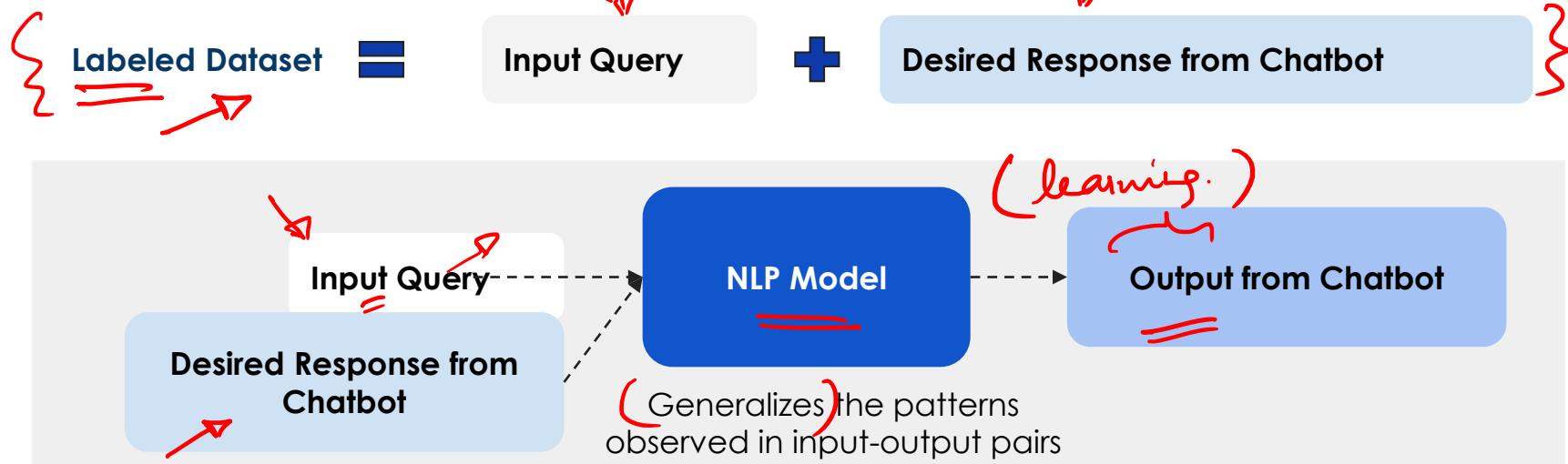


This file is meant for personal use by amitava.basu@gmail.com only.

Proprietary content. ©Great Learning. All Rights Reserved. Unauthorized use or distribution prohibited

Training & Tuning

Model Training



+ → **Rule based Chatbots**

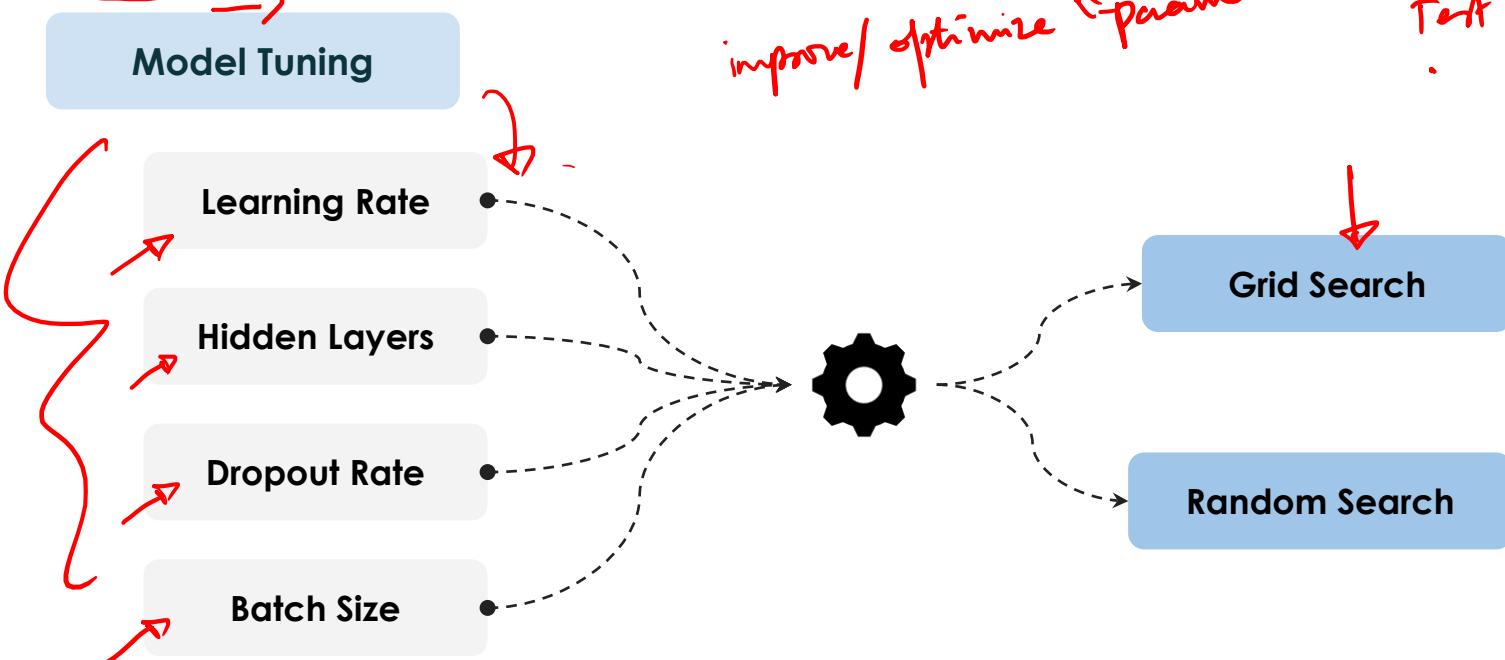
→ **Generative Chatbots**

× → **Retrieval Chatbots**

Training & Tuning

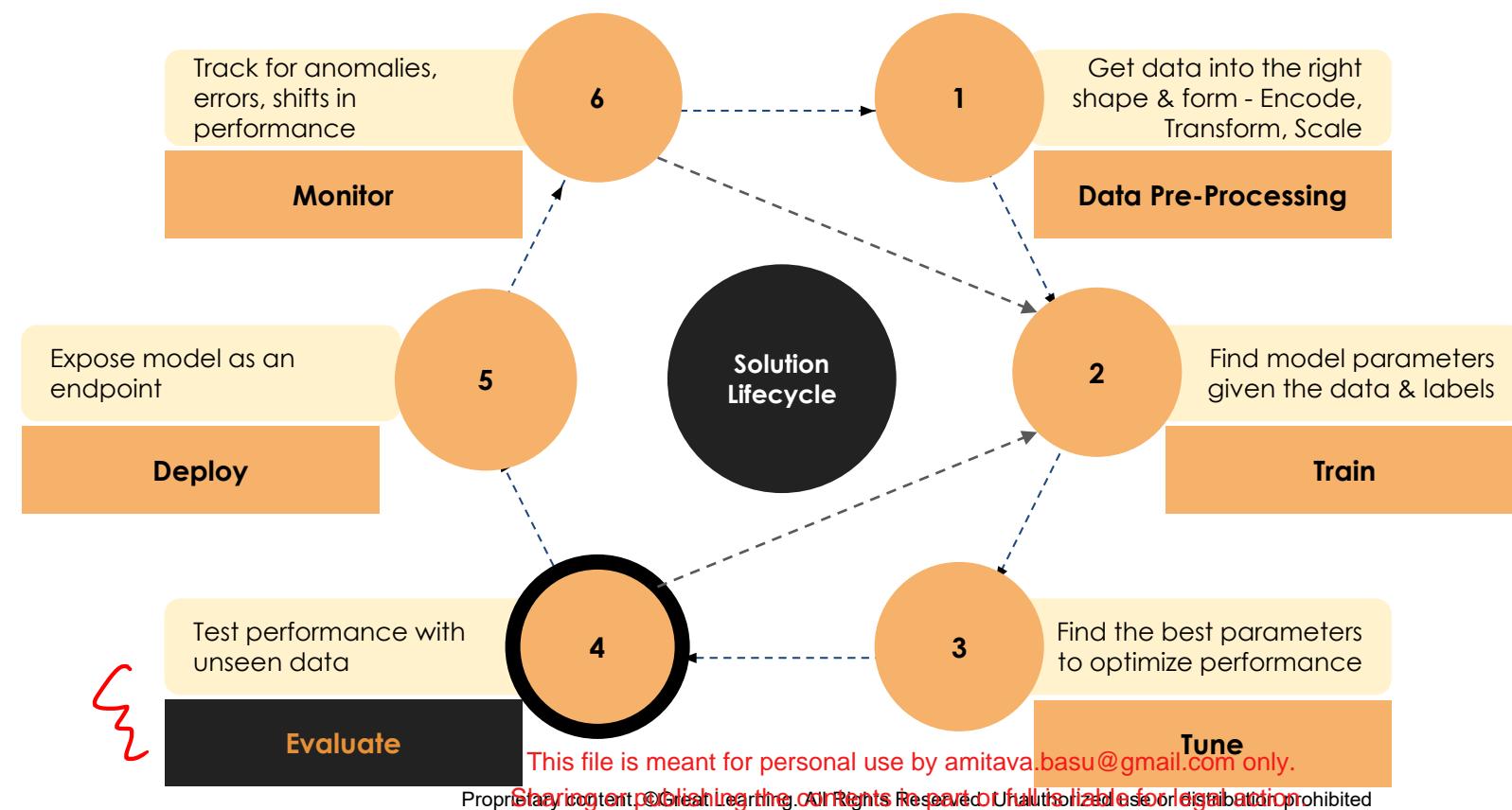
"Hyperparameters control the behavior of the model during training."

improve/ optimize "parameters!"

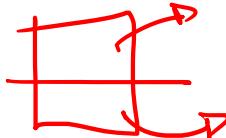


Hyperparameter tuning involves testing different combinations of hyperparameters and selecting the combination that offers the best performance.

A Typical Solution Lifecycle



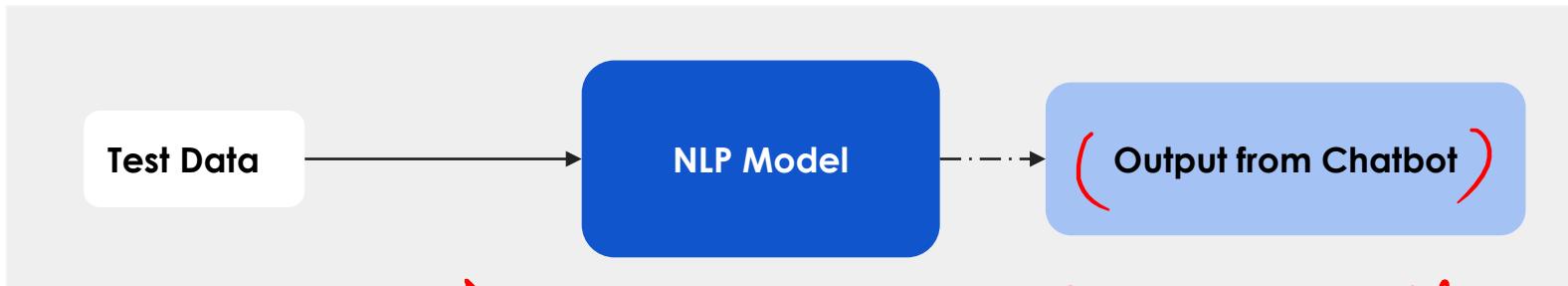
Performance Evaluation



↔ ~ pattern..

Model Testing

Using Validation/Test Dataset that is different from training data.



Test Data

NLP Model

Output from Chatbot

Evaluation

Human

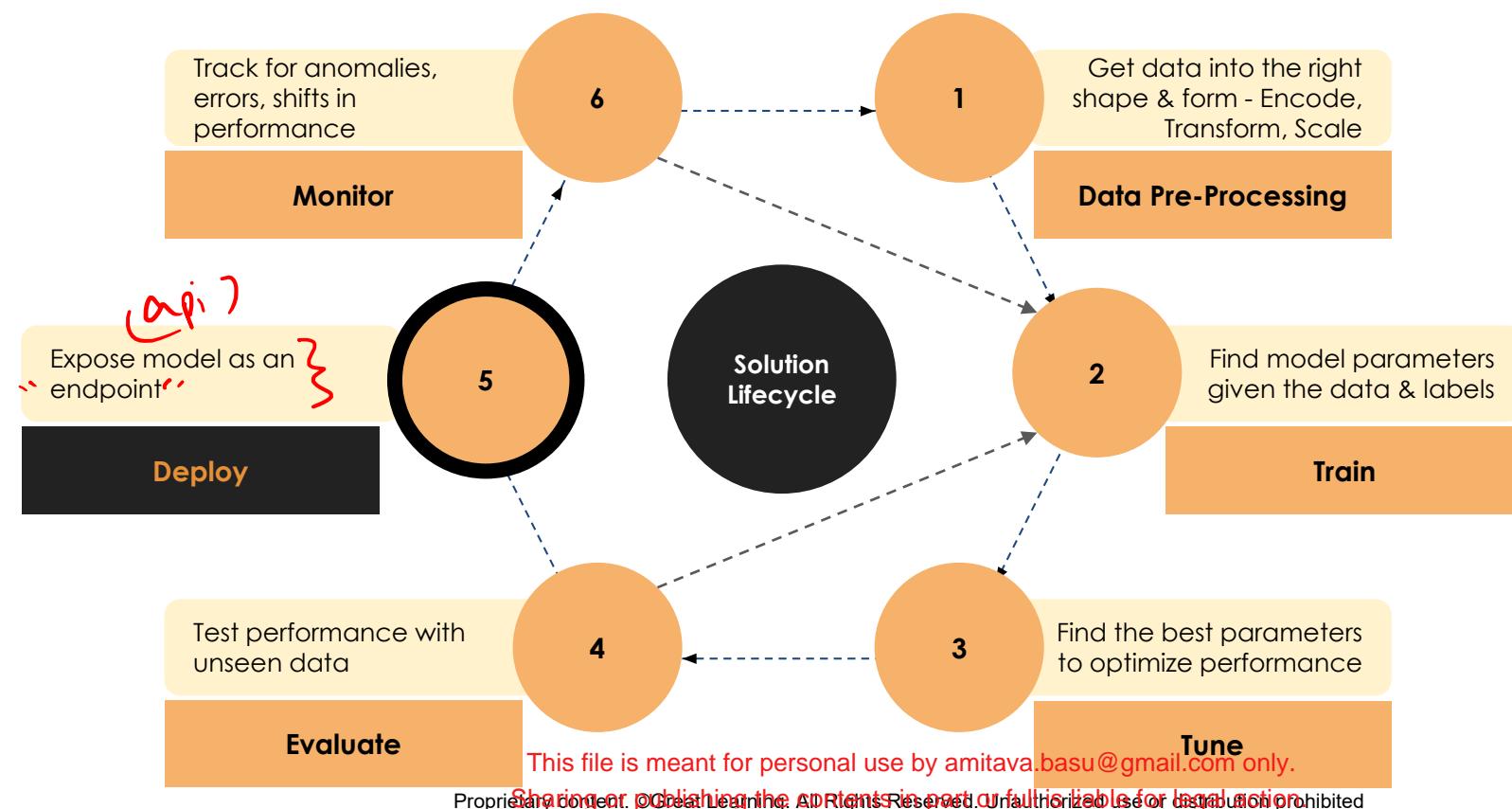
Validate & label

Generalize

Well on new examples

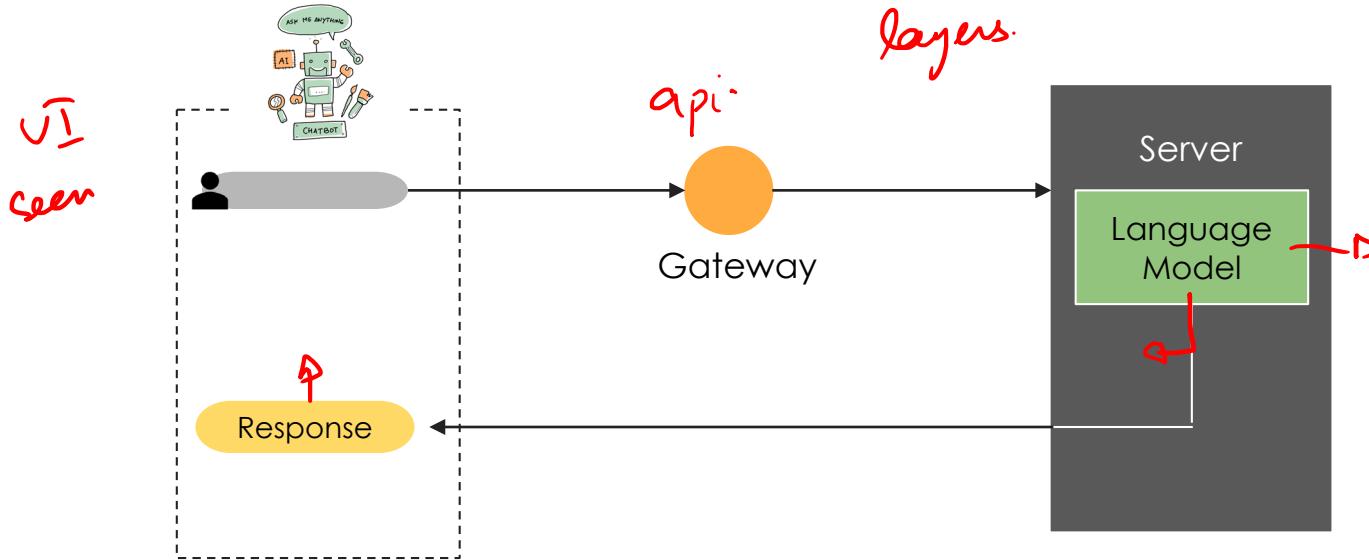
Sept'21 → model

A Typical Solution Lifecycle

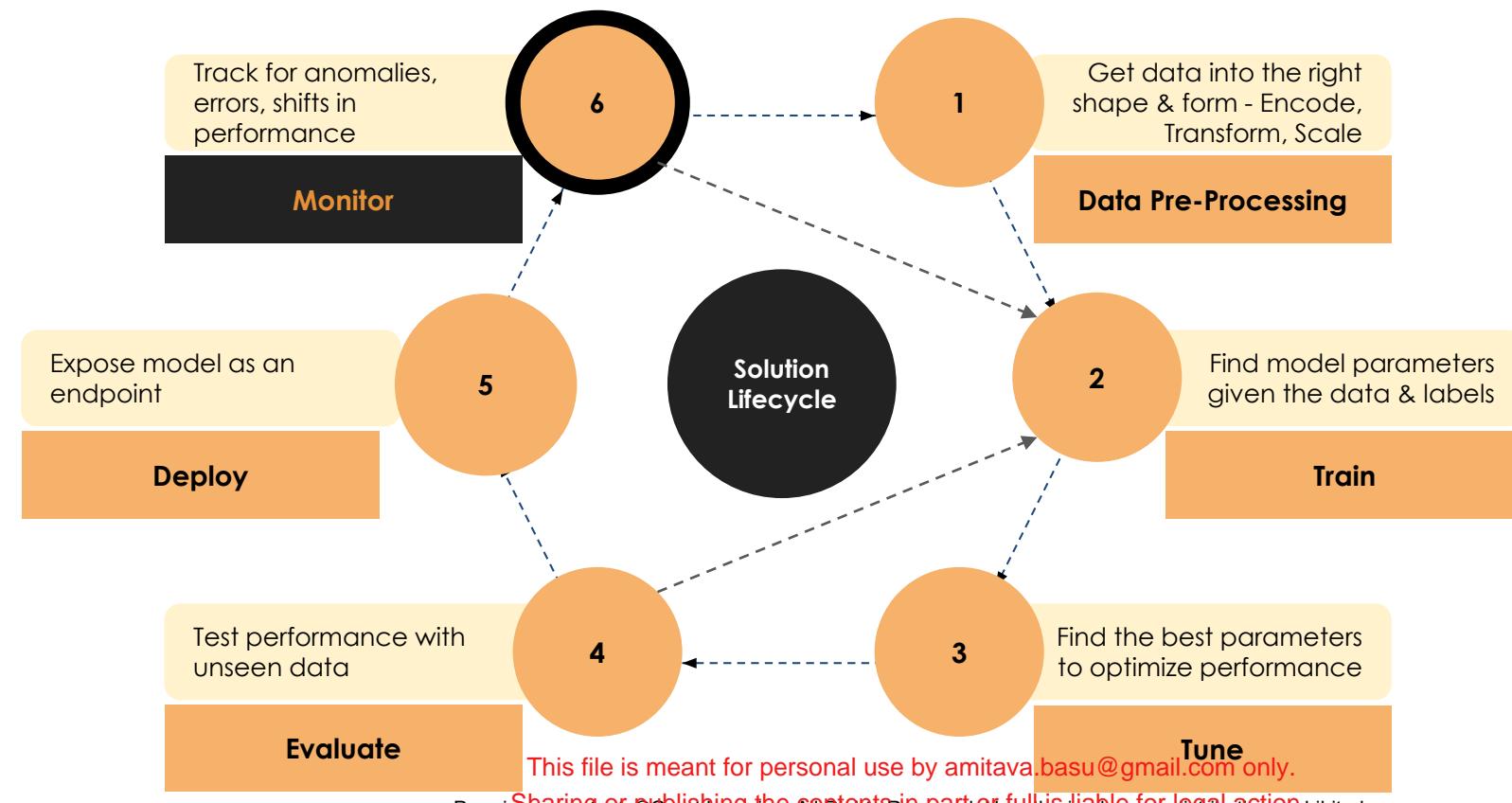


Model Deployment

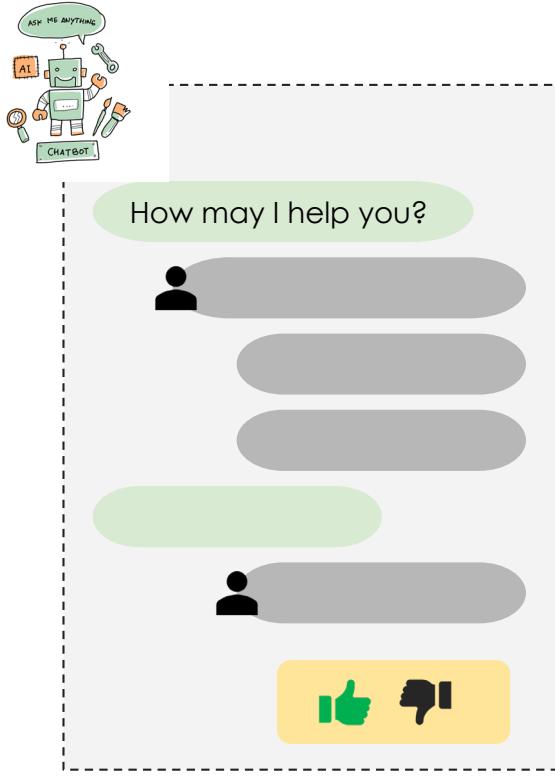
“How am I able to get the real time response for my queries?”



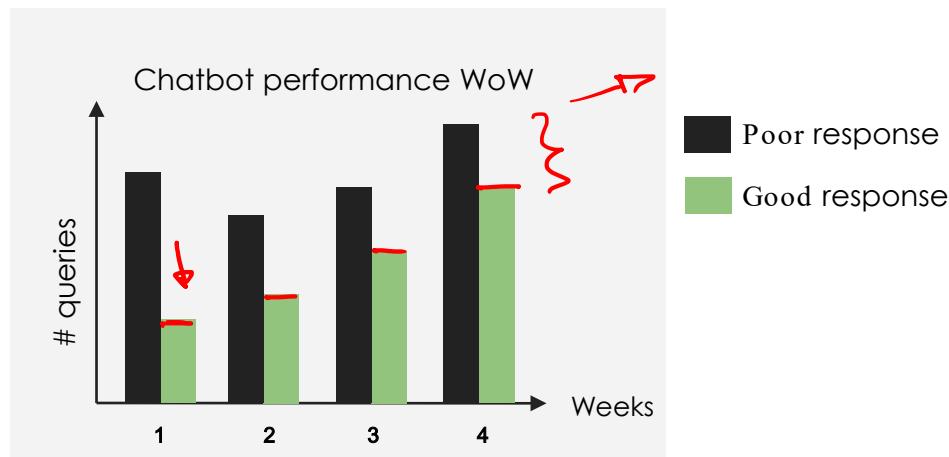
A Typical Solution Lifecycle



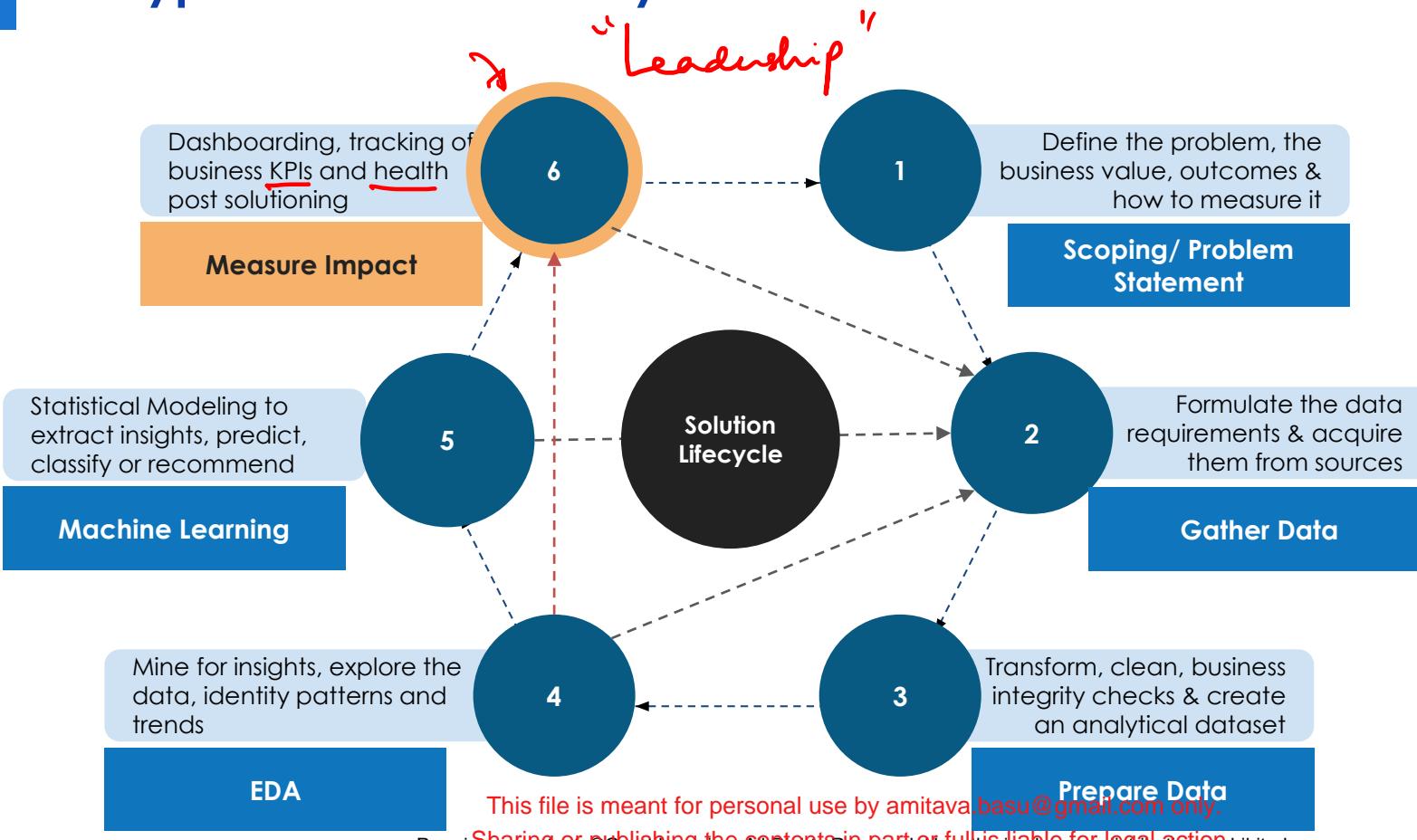
Model Monitoring



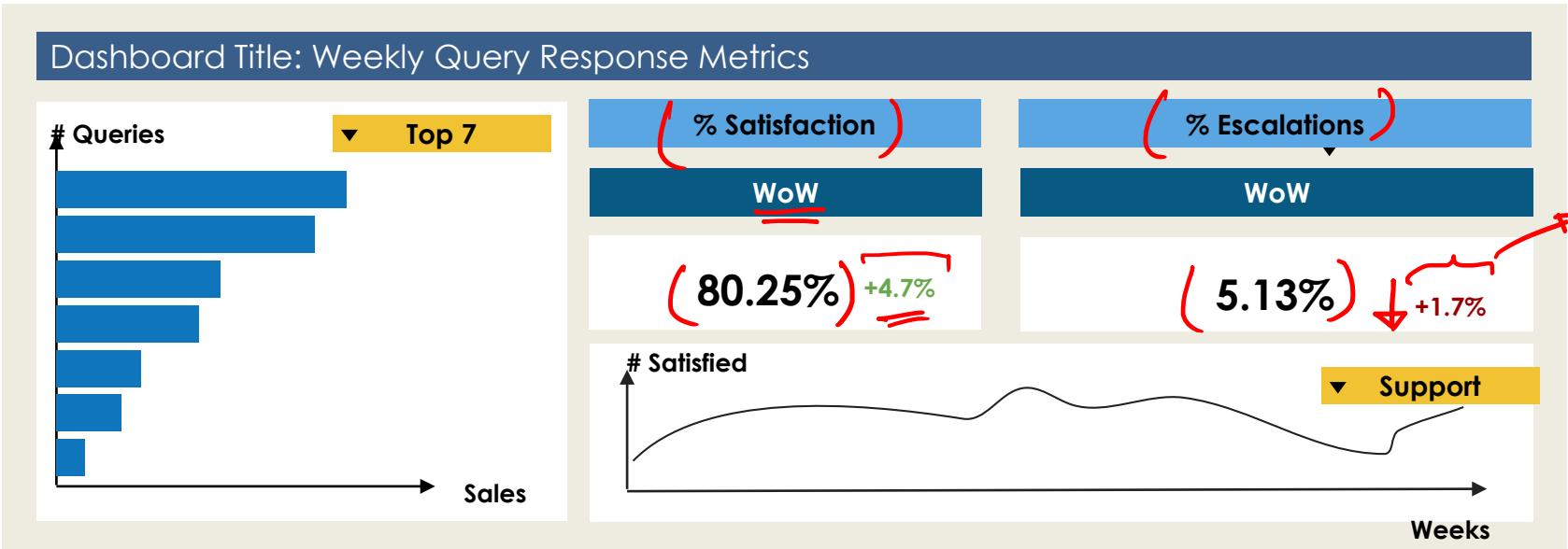
Date	C-ID	Category	Query	Response	Satisfied
					✓
					✗
					✓



A Typical Solution Lifecycle



Measuring Impact



Summary

ART \Rightarrow interpretation - comm - decisionmaking
"judging"

1. Data science lifecycle is a cycle needing collaboration across teams and stakeholders
2. ML lifecycle is contained within Data Science lifecycle
3. Art is to define problems, get the right data, take right decisions during the lifecycle and the right final recommendations for deployment and improvement
4. Science is to be able to code the algorithms, extract patterns & insights and deploy the solutions



Happy Learning !

