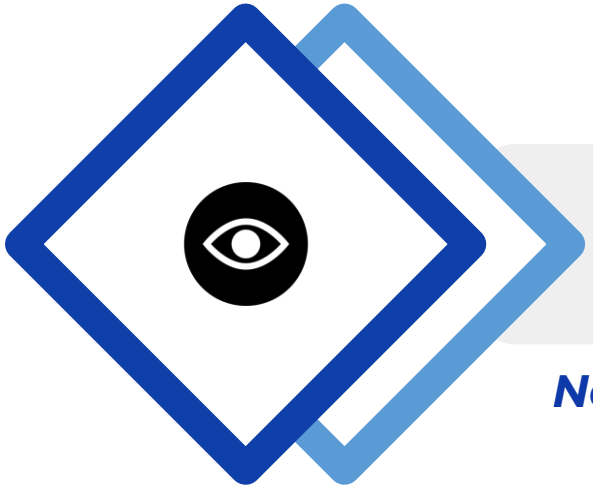


Data Science Lifecycle

Navigating the Data Science Lifecycle

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

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

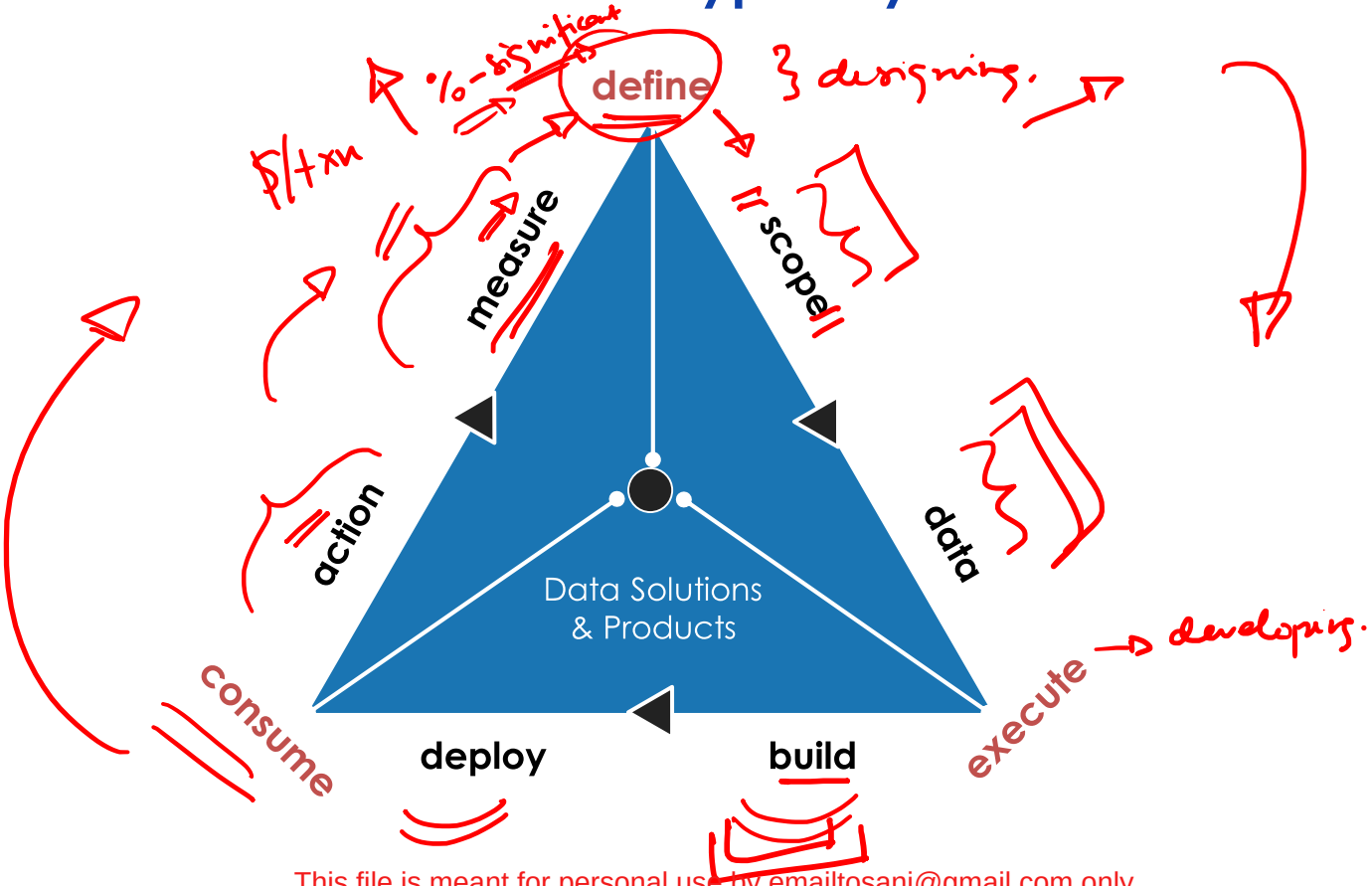
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Agenda

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

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

How are Data-driven solutions typically built?



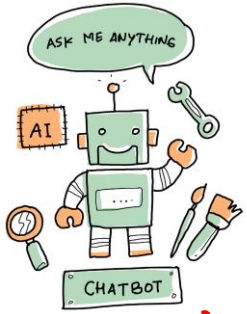
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Business Context - E-commerce

- Product selection struggle
- Difficult navigation of web page
- Inadequate support

- Poor shopping experience
- Poor customer satisfaction
- High cart abandonment rate

- Track orders



- Search for products

- 24*7 support

- Assist with FAQs

- Product analytics

- Nudge to complete sales

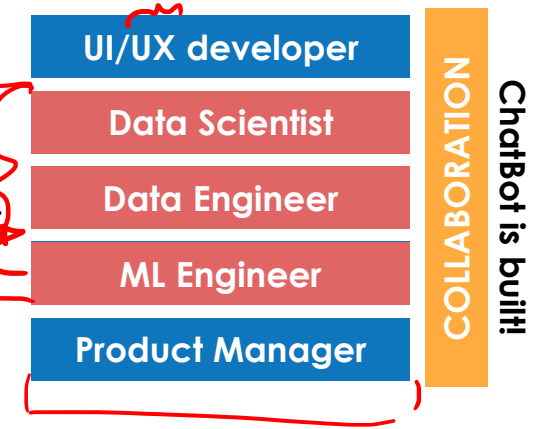
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Strategy for building the ChatBot

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

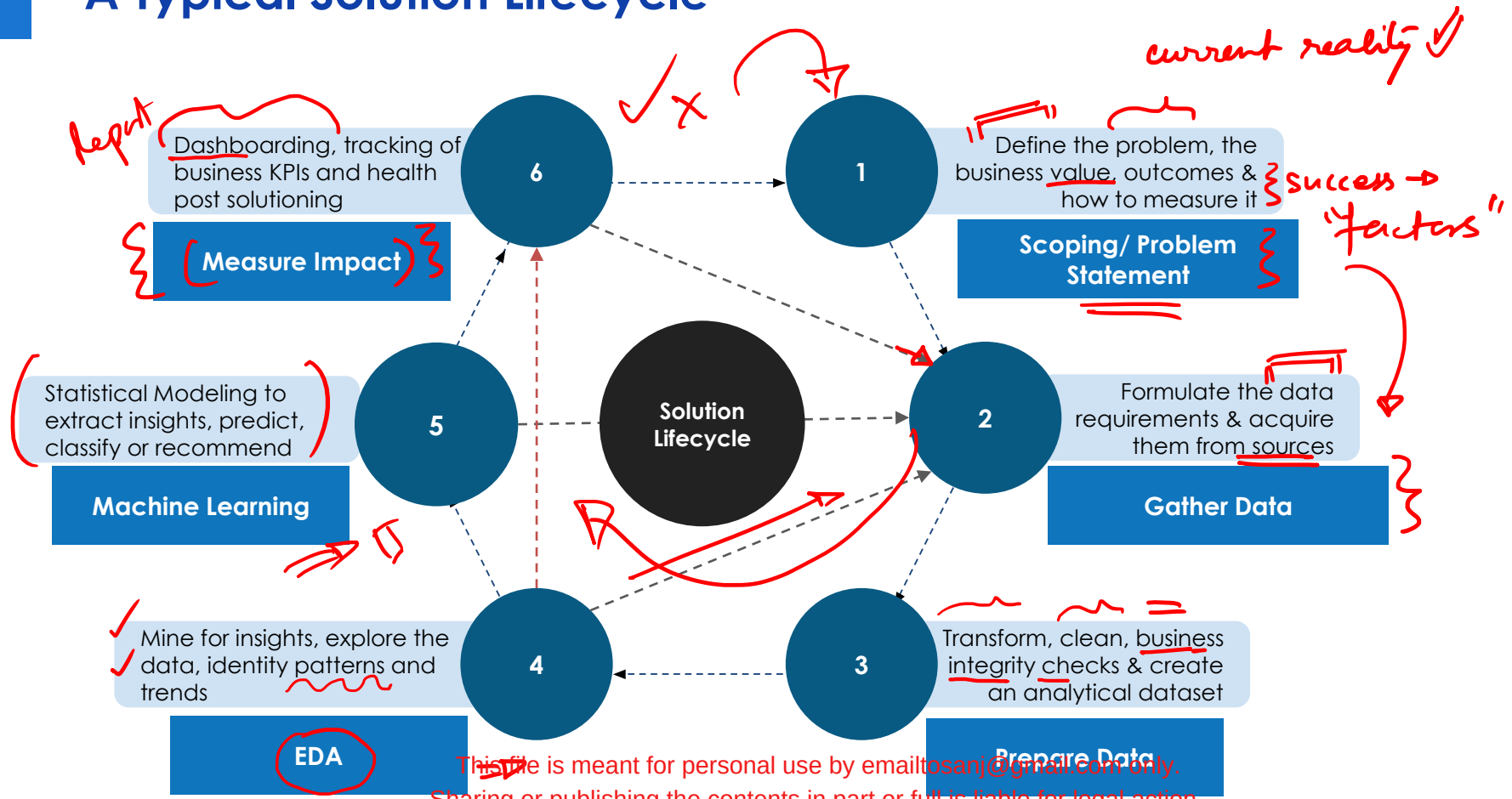
You would need someone to -

- 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 it's performance.
- Manage and drive the development & consumption of this product.



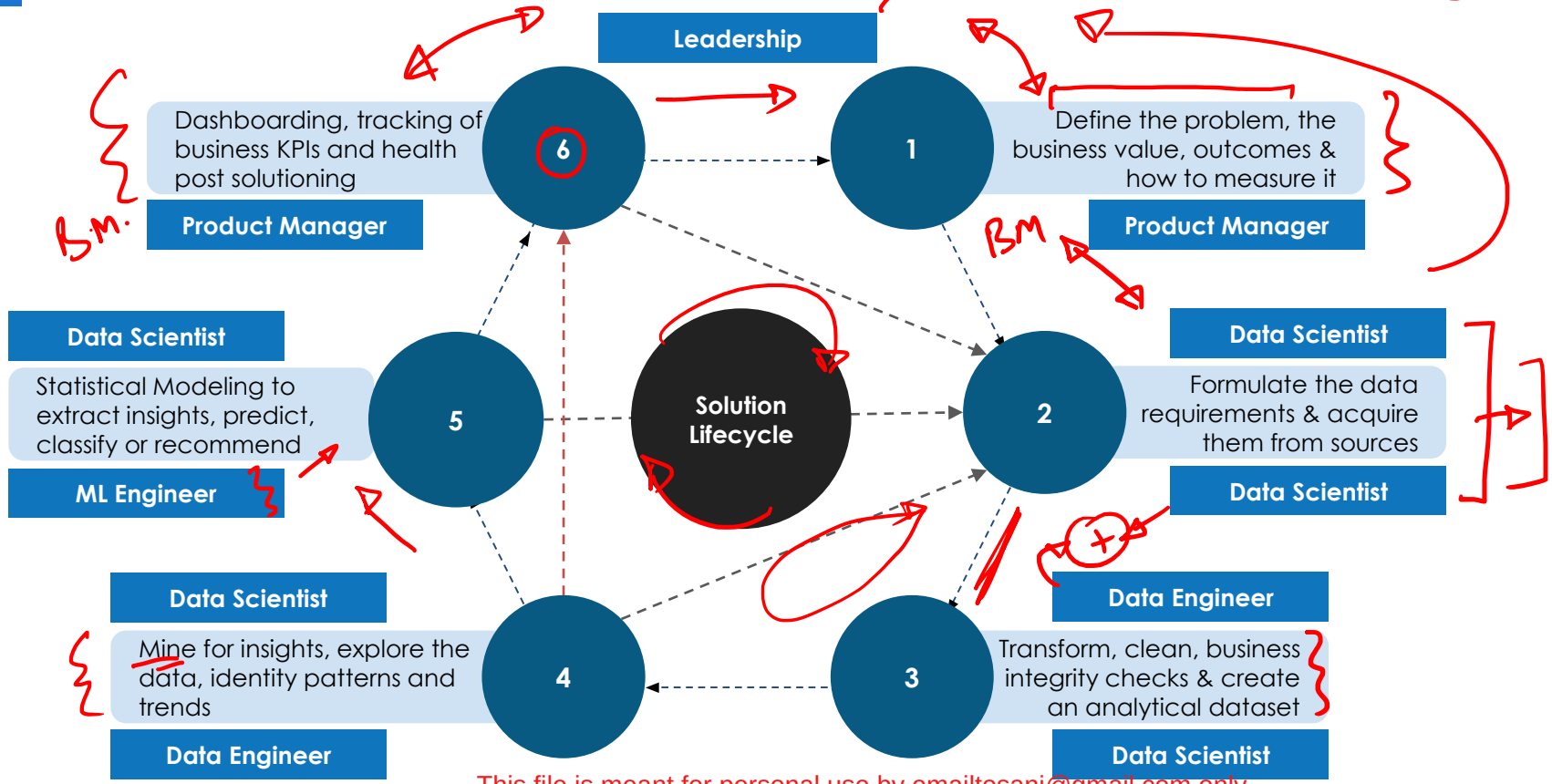
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

A Typical Solution Lifecycle



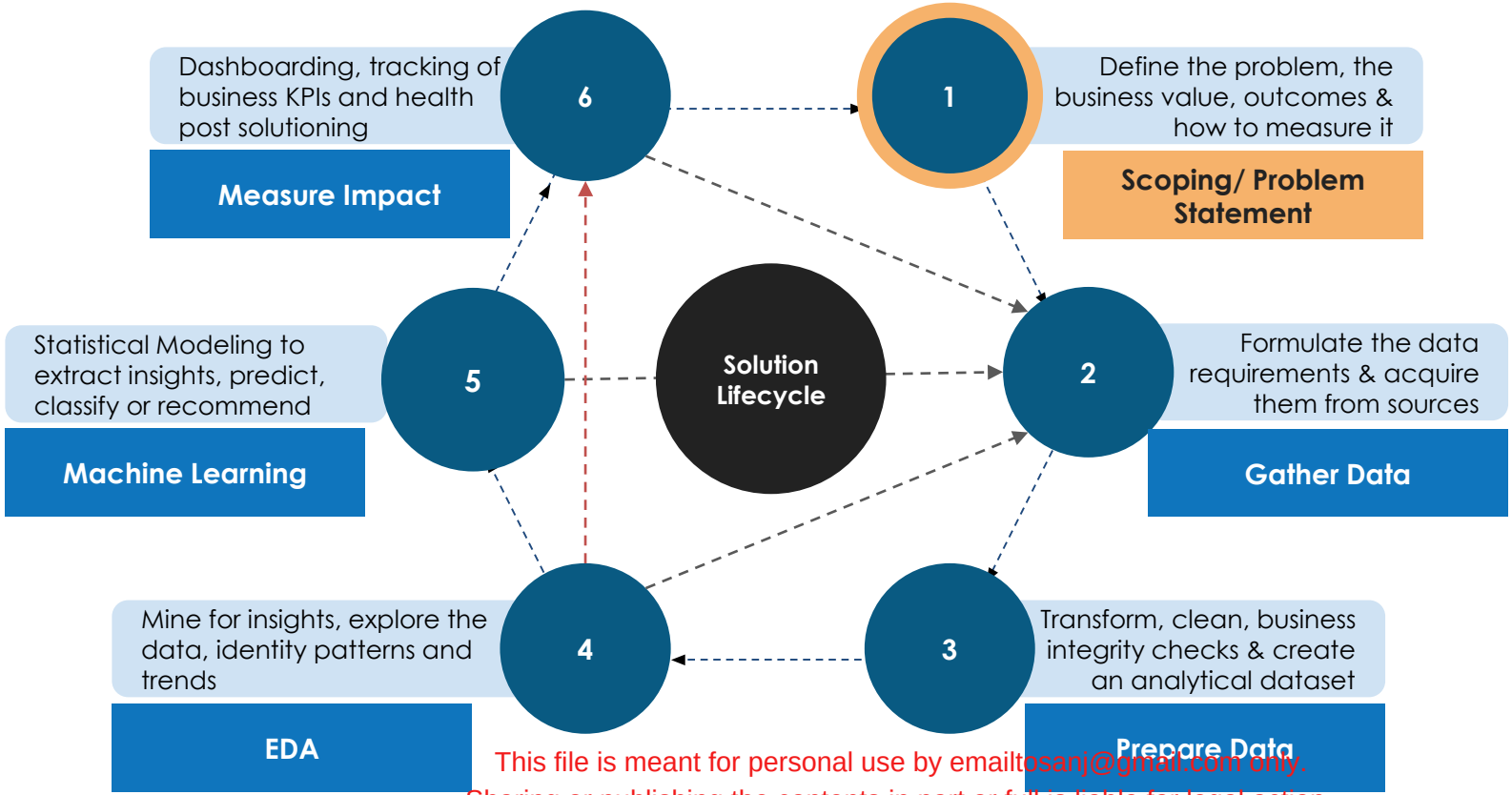
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

A Typical Solution Lifecycle



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

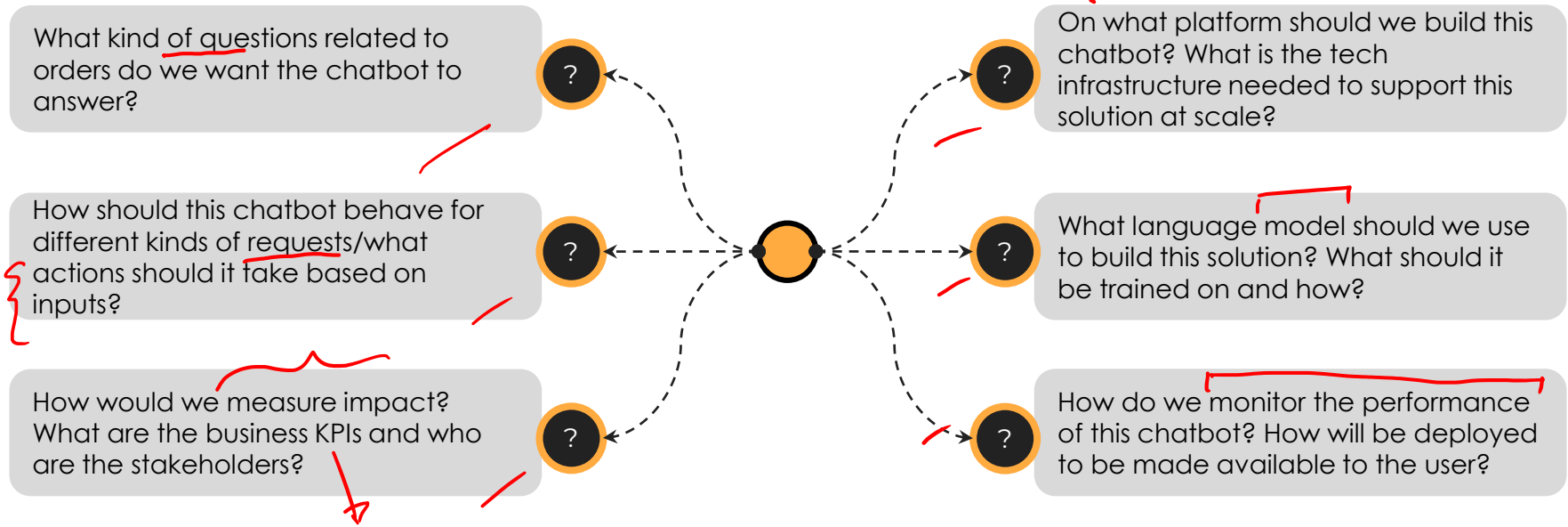
A Typical Solution Lifecycle



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

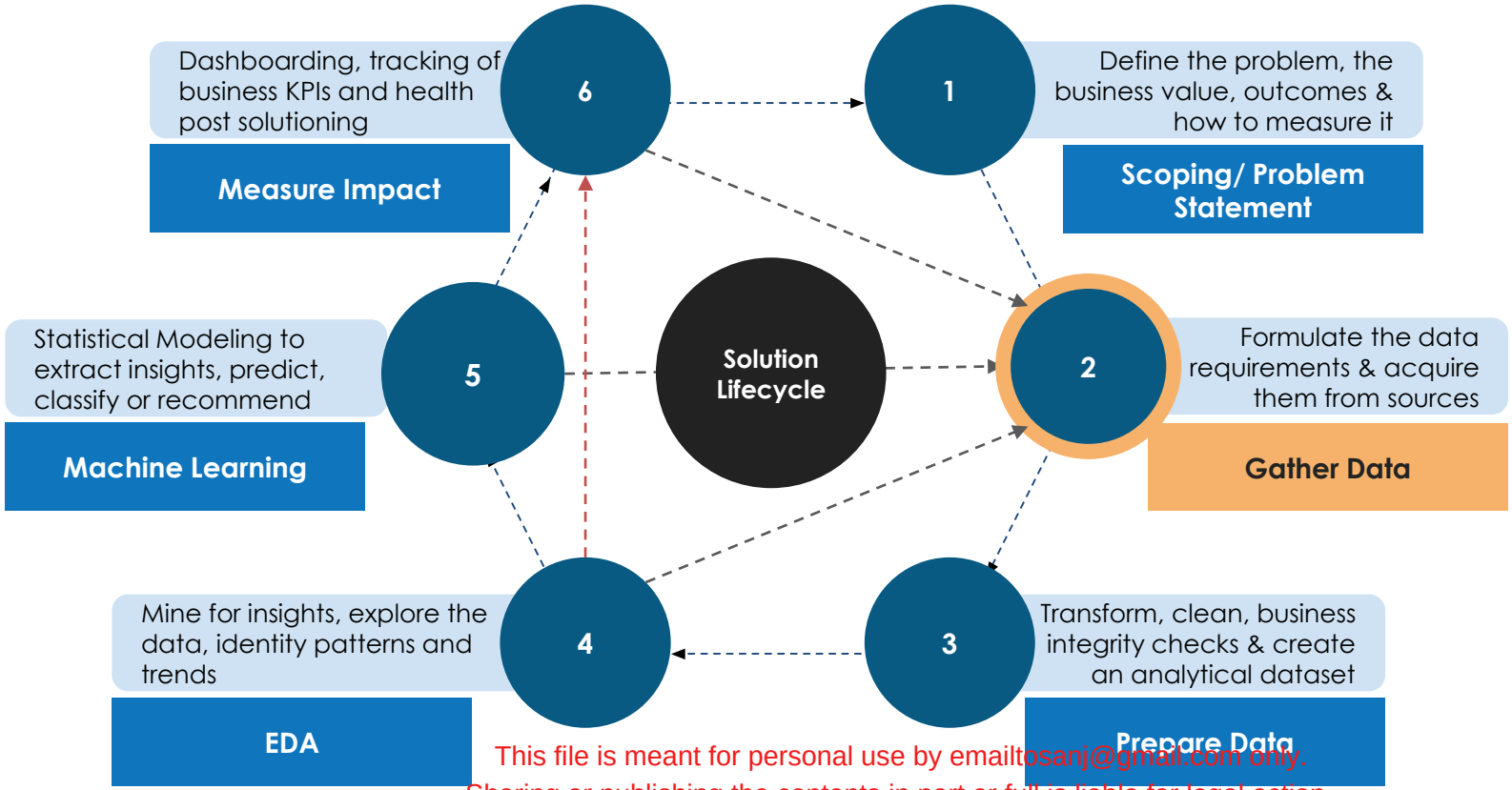
Scoping/Problem Definition

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



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

A Typical Solution Lifecycle



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

Gather Data

"What data do I need to build this chatbot?"

User data

- Customer details
- Browsing
- Purchase History

Product data

- Product Brand
- Product type
- Model type

Orders data

- Order ID
- Payment Type
- Order quantity

- Payment Methods
- Product availability
- Refund/Return

FAQ

- Product quality
- Service rating
- Delivery rating

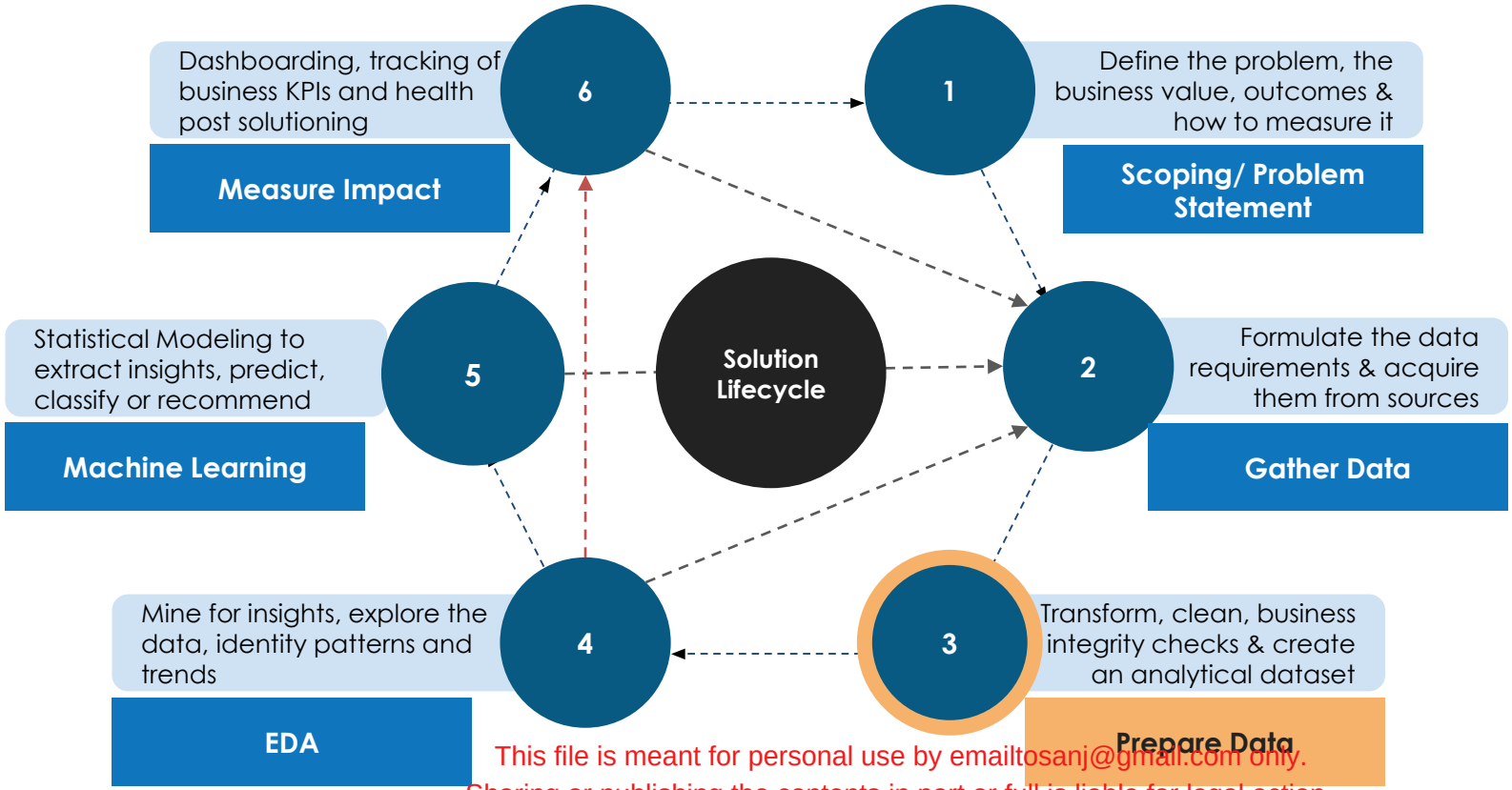
Feedback

- Pending orders
- Dispatched orders
- Returned orders

Shipping

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

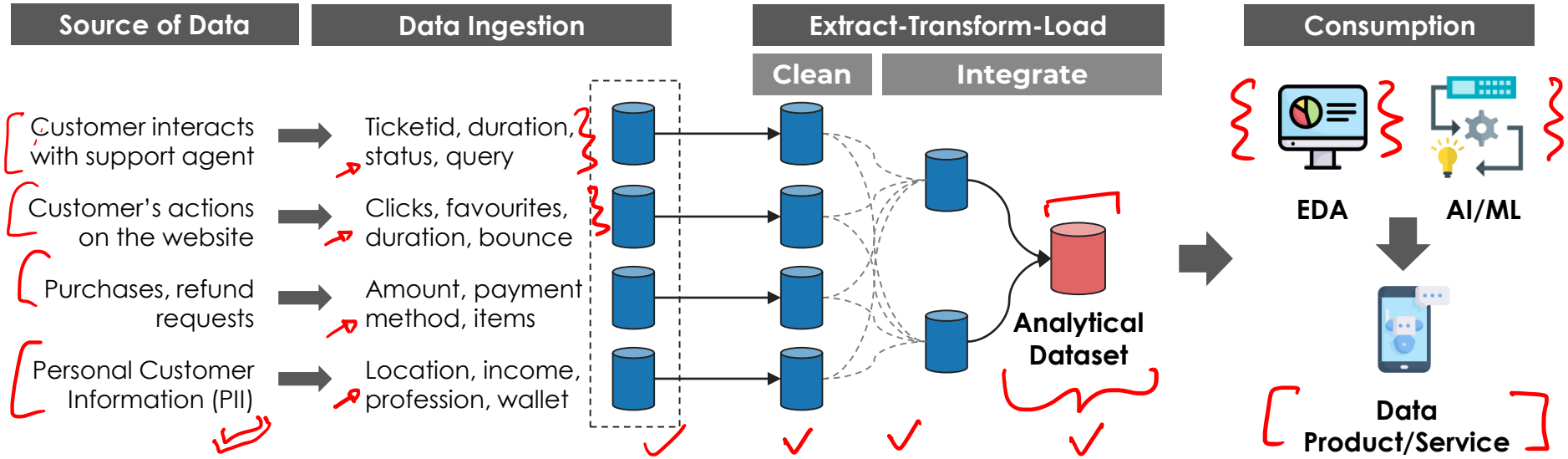
A Typical Solution Lifecycle



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

Prepare Data

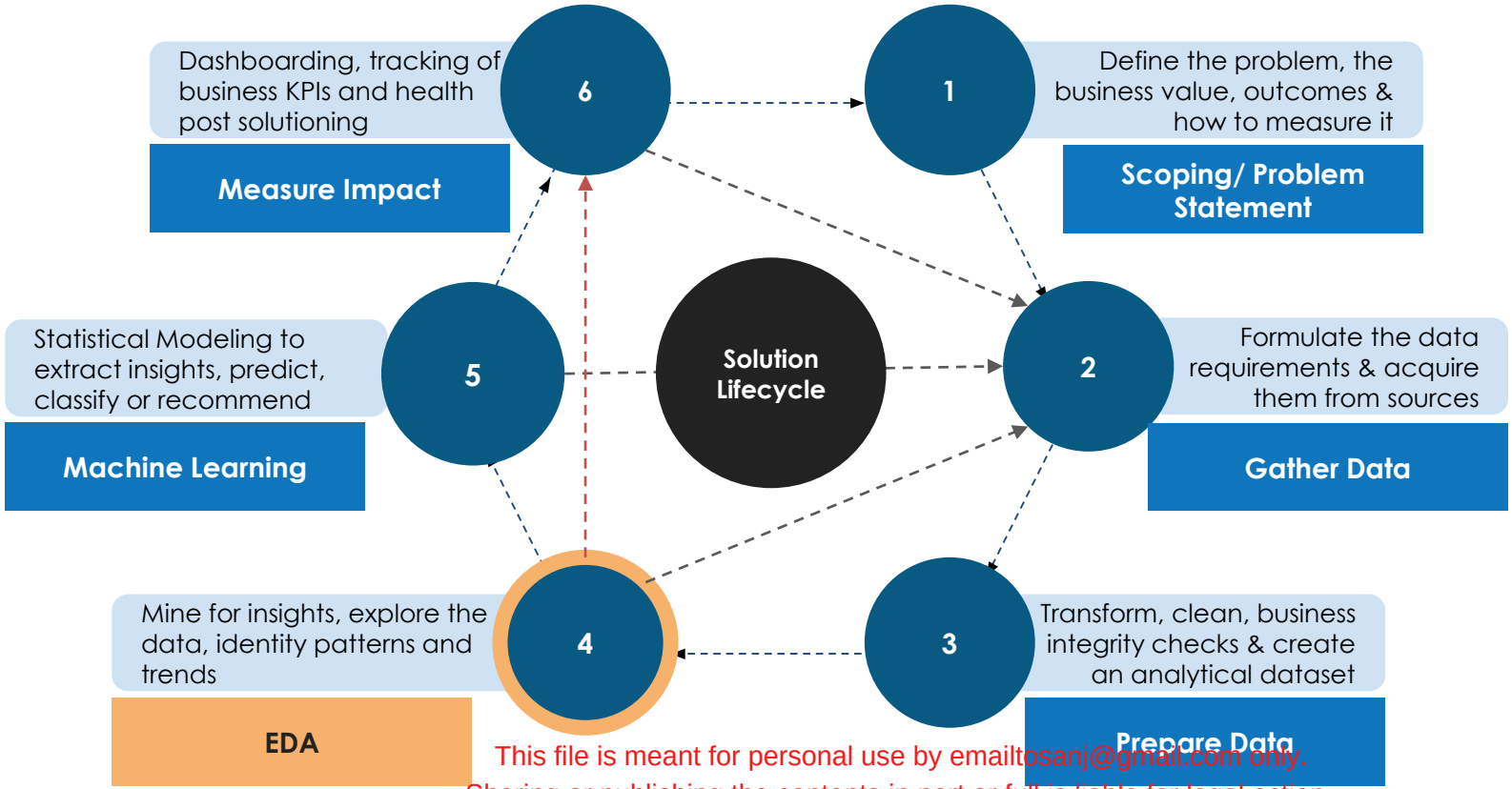
"How do I create the data I need for my data solution/product?"



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

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

A Typical Solution Lifecycle



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

Exploratory Data Analysis

"How should I approach exploring my data?"

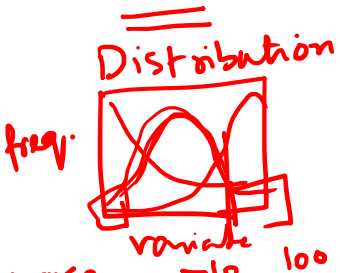
- What is the user's intent?
- What product/service is being referred to?
- What are the data sources for this query?

- Common pain points for customers
- Feedback on chatbot performance
- Key product/service features

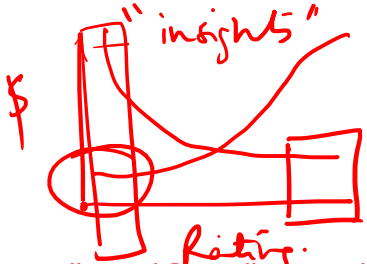
Data Cleaning → **Formulate Questions** → **Data Analysis** → **Conclusions** → **Documenting Insights**

- Removing/ treating missing values
- Treating outliers
- Variable conversion

Univariate Analysis



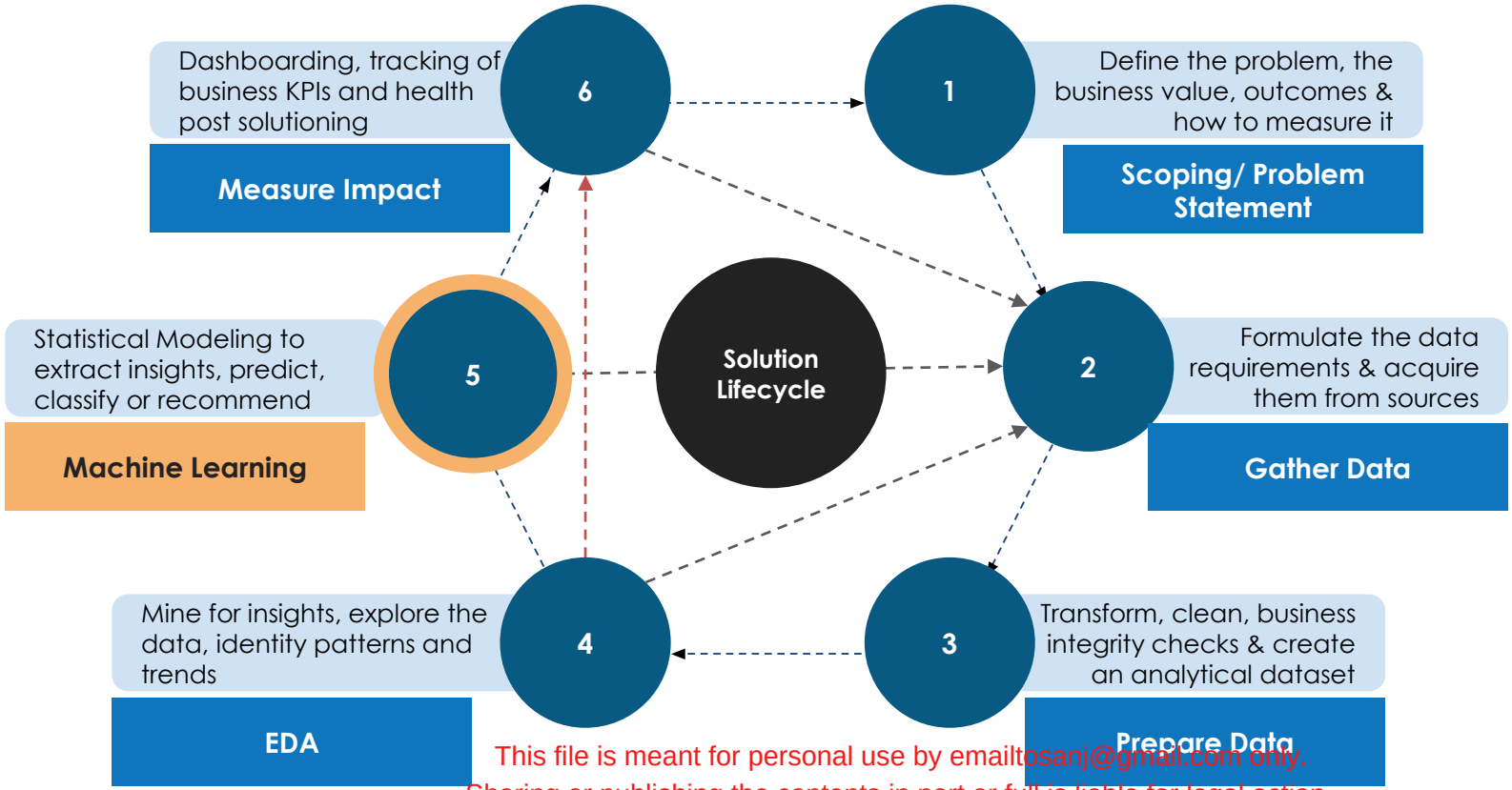
Bi/multivariate Analysis



- The most frequent services opted for in each category
- Improvements to make based on the findings

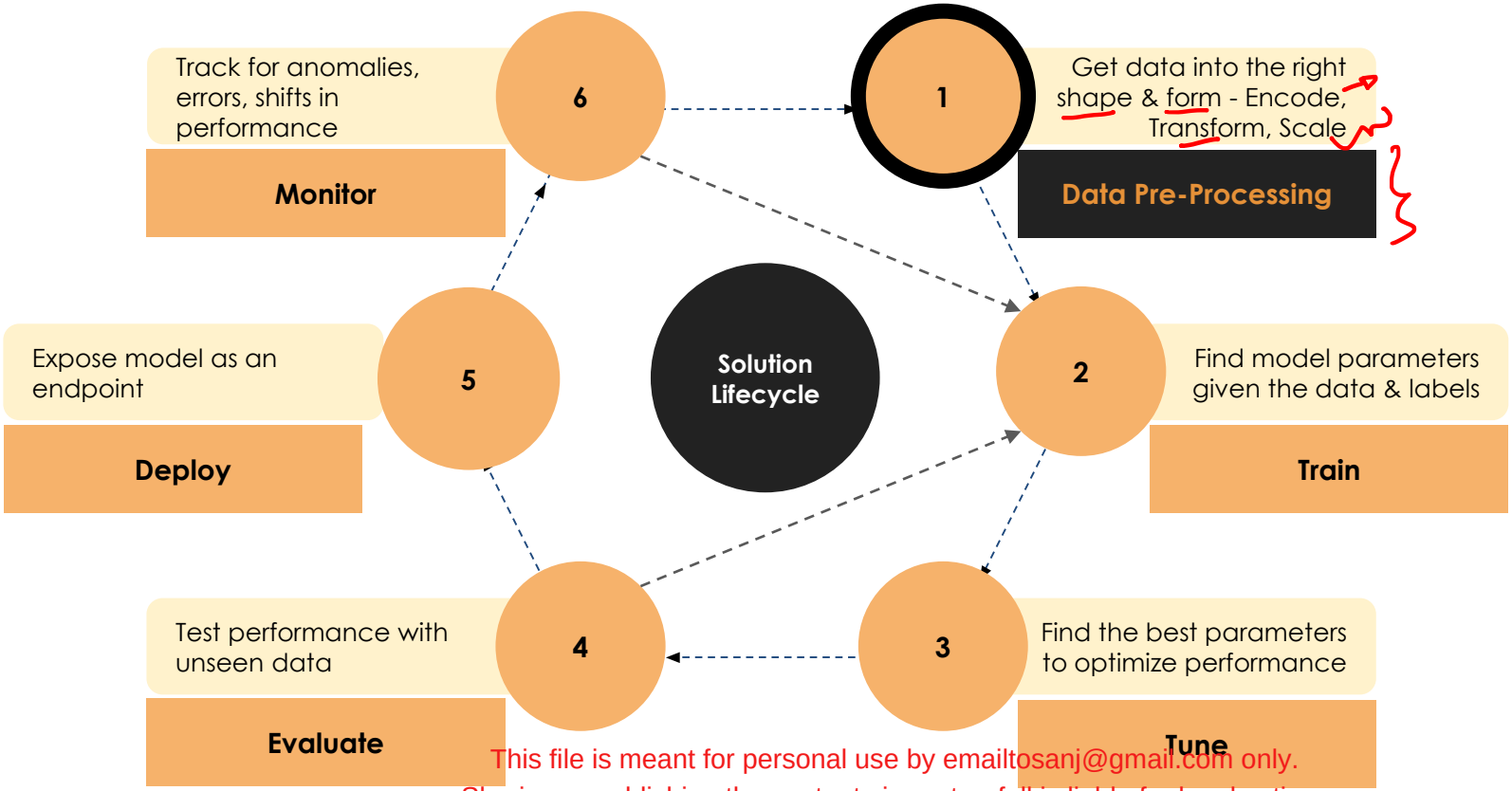
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

A Typical Solution Lifecycle



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

A Typical Solution Lifecycle



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

Data Pre-processing

	Input	Input type
User Input to chatbot →	When will my order be delivered ?	Order Status

How does the chatbot understand us?

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

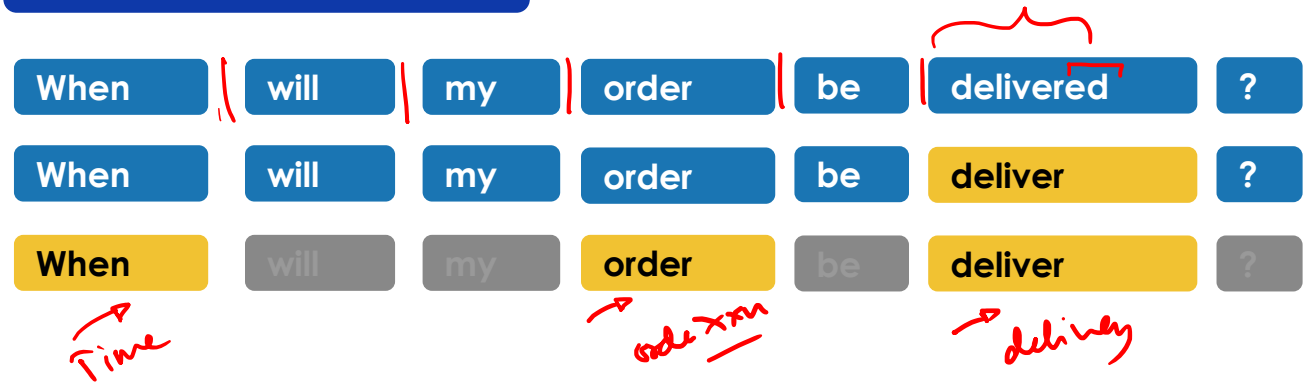
This is how a machine understands - This is a customer, who seems to have an overall positive sentiment, and wants to purchase a new HP laptop.

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

Data Pre-processing

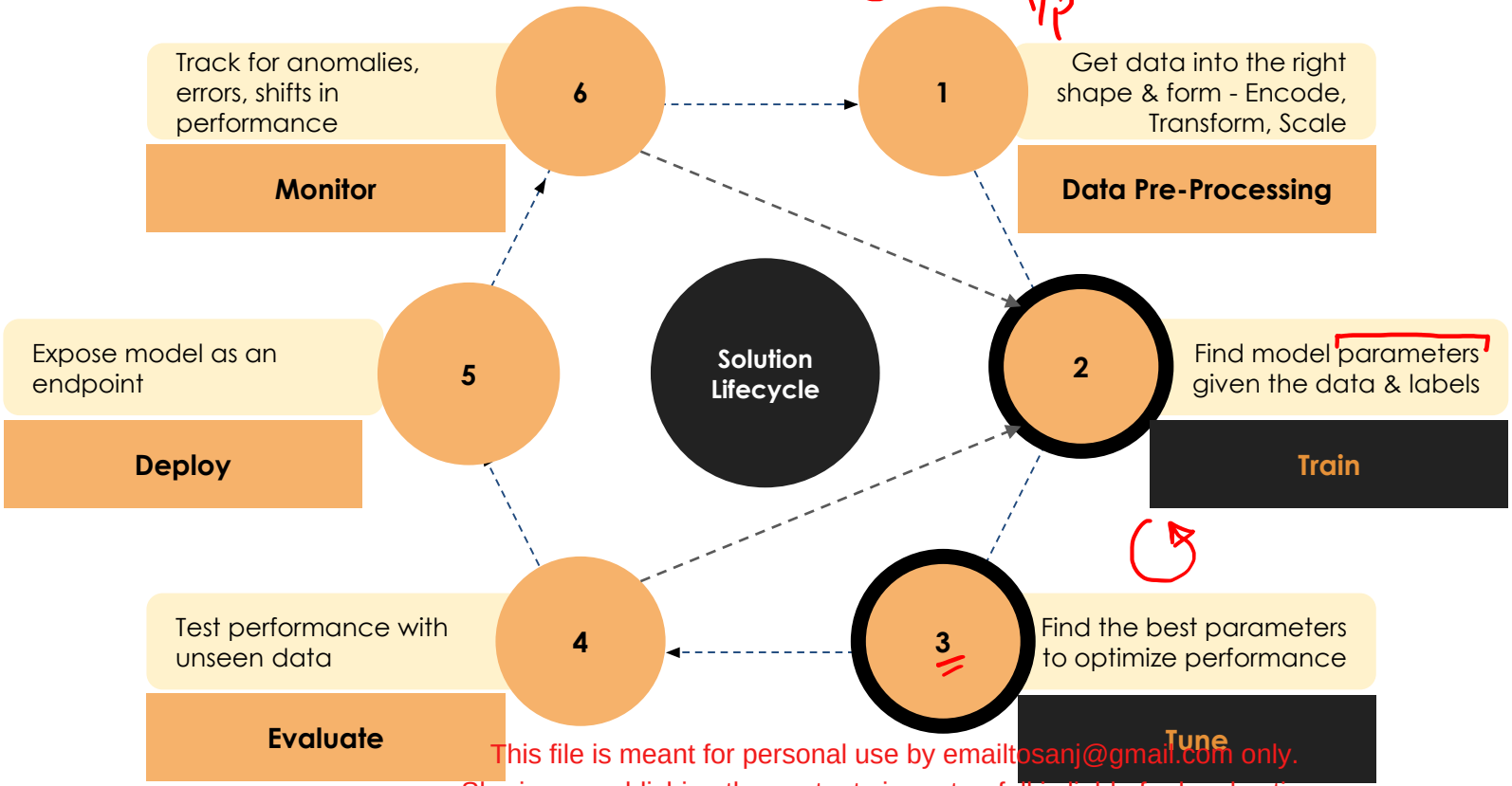
	Input	Input type
User Input to chatbot →	When will my order be delivered ?	Order Status

Data Pre-processing steps



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

A Typical Solution Lifecycle

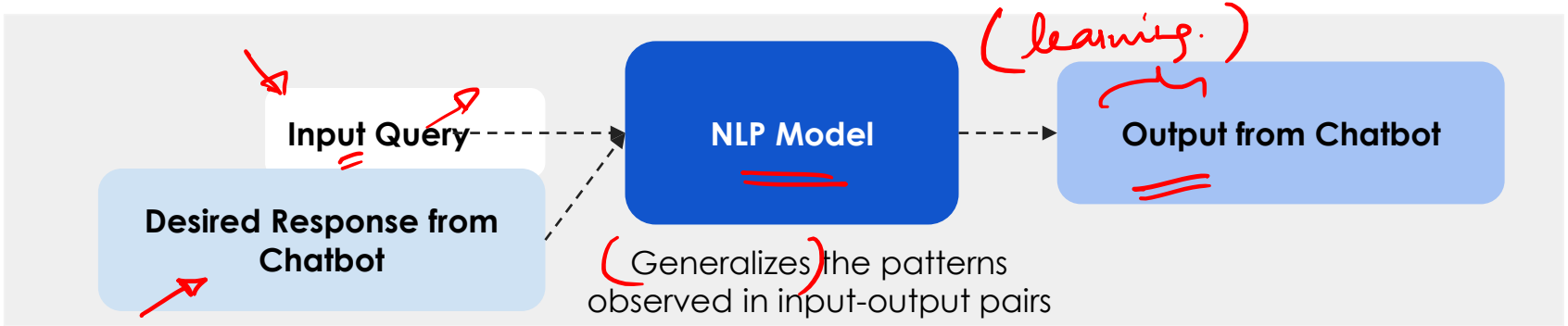


This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Training & Tuning

Model Training

Labeled Dataset = Input Query + Desired Response from Chatbot

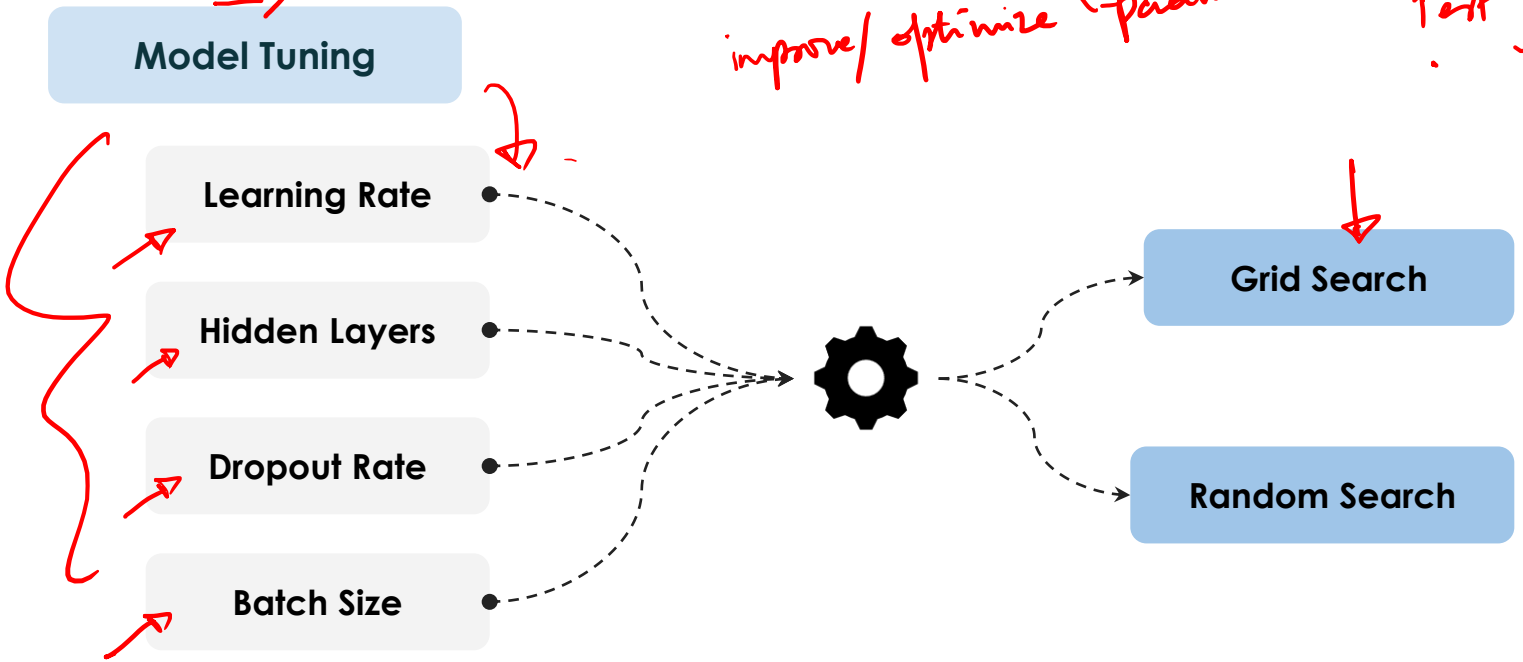


- Rule based Chatbots
- Generative Chatbots
- Retrieval Chatbots

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

Training & Tuning

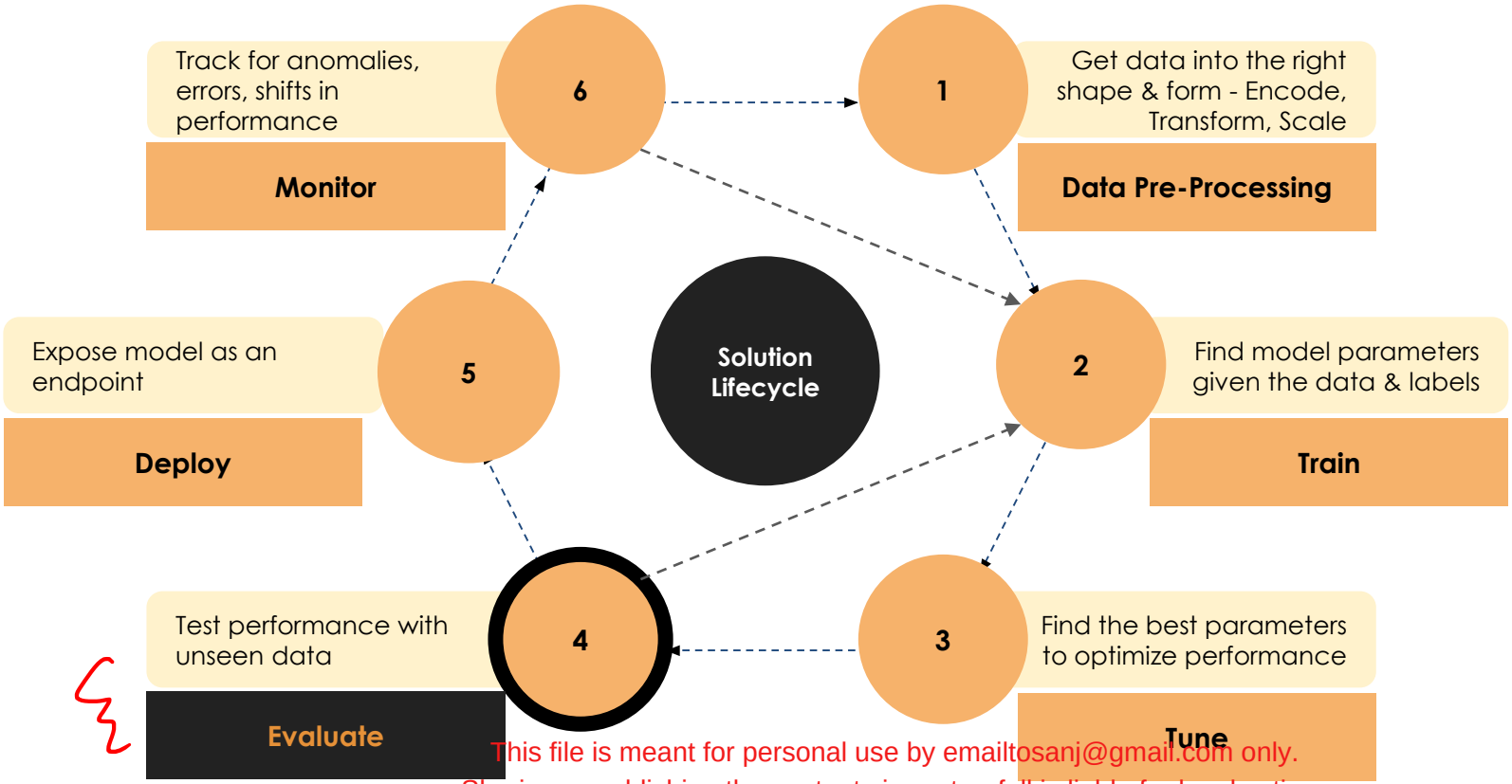
"Hyperparameters control the behavior of the model during training."



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

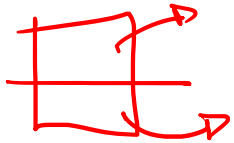
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

A Typical Solution Lifecycle

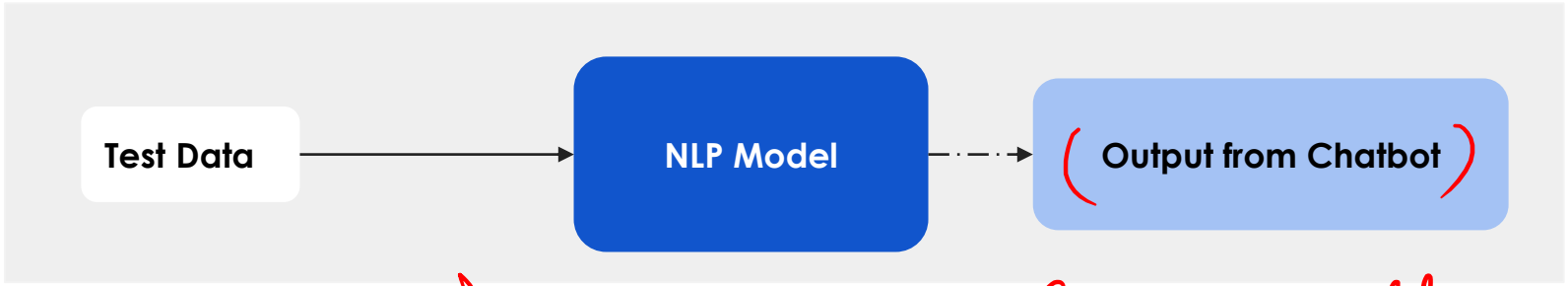


This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Performance Evaluation



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

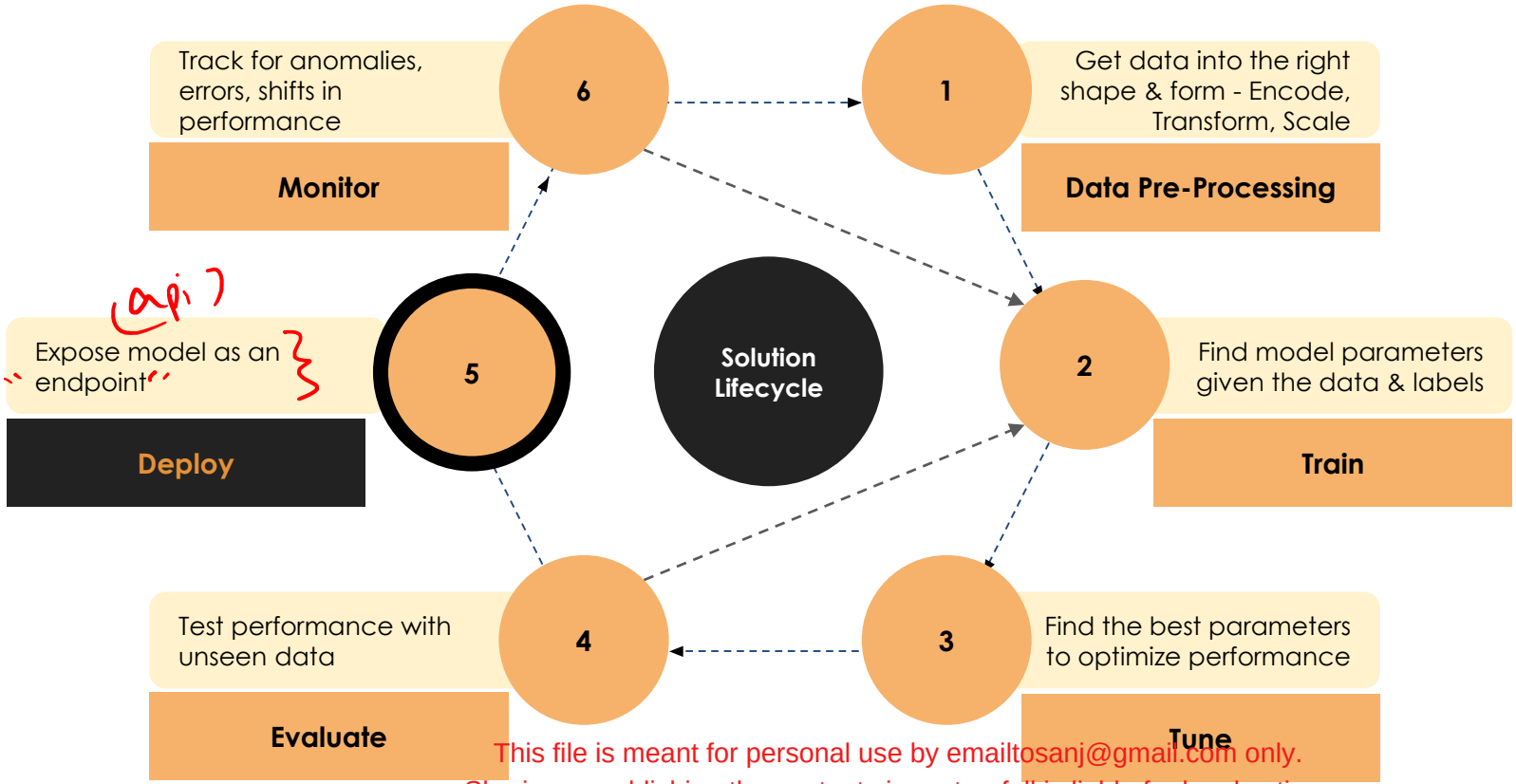


Evaluation

- **Human** — Validate & label
- **Generalize** — Well on new examples

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

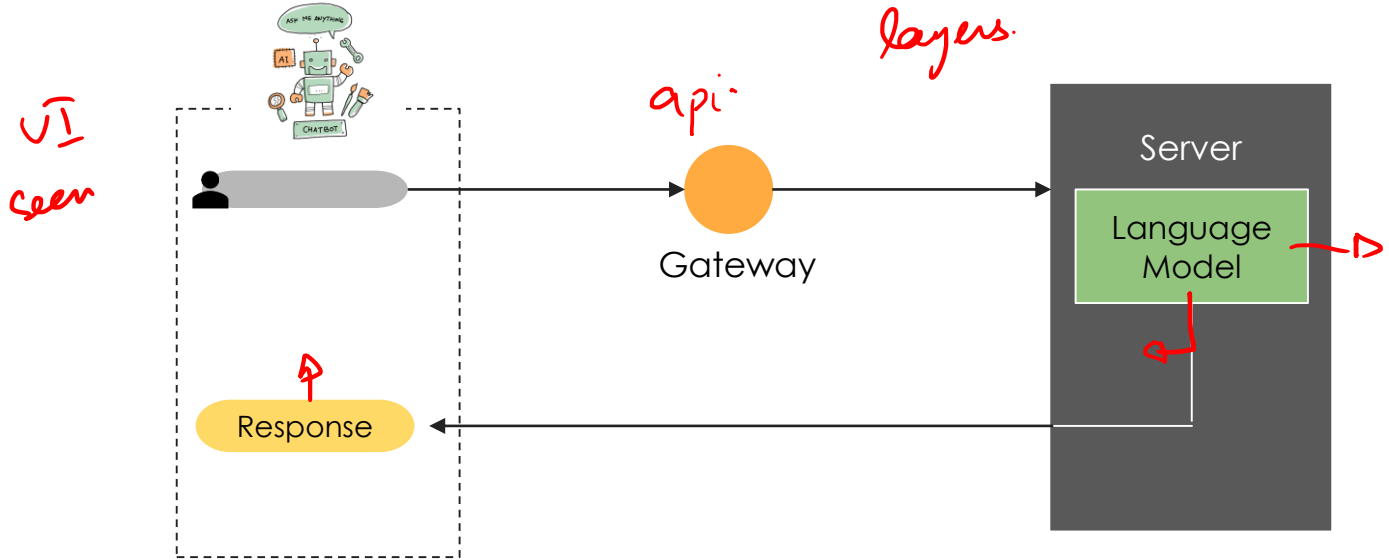
A Typical Solution Lifecycle



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

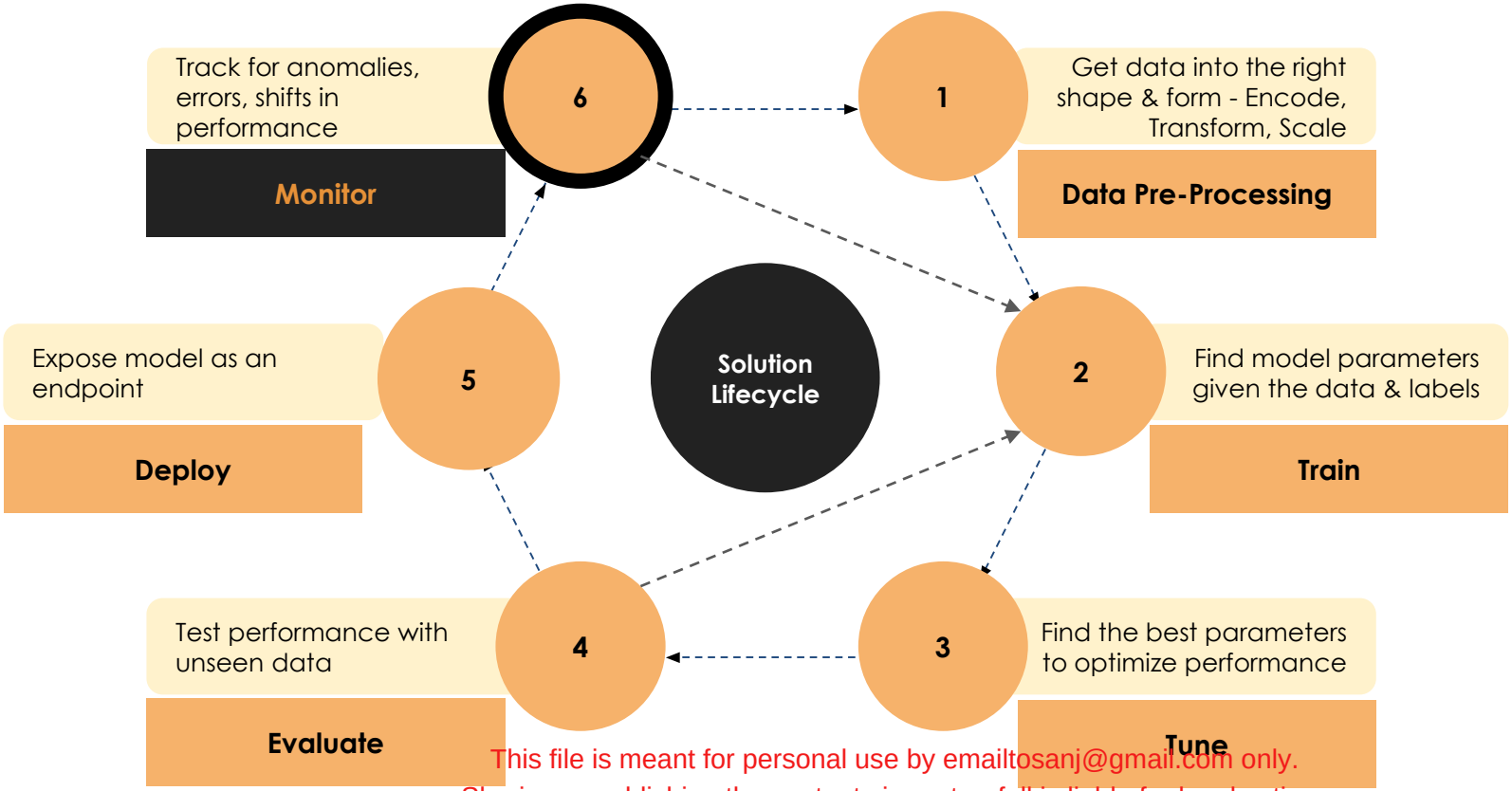
Model Deployment

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



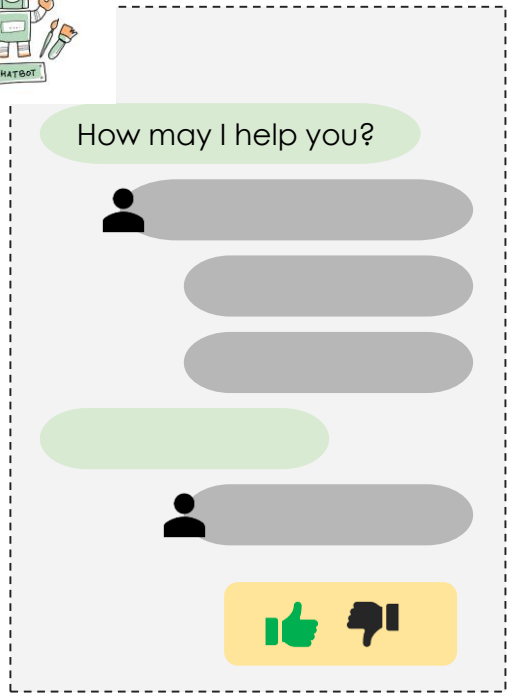
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

A Typical Solution Lifecycle

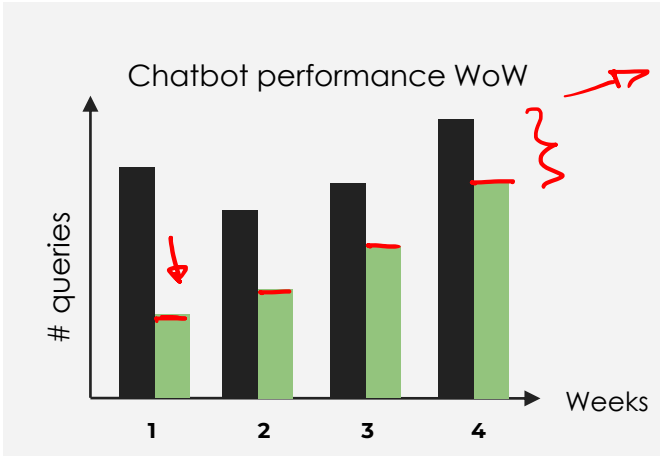


This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

Model Monitoring

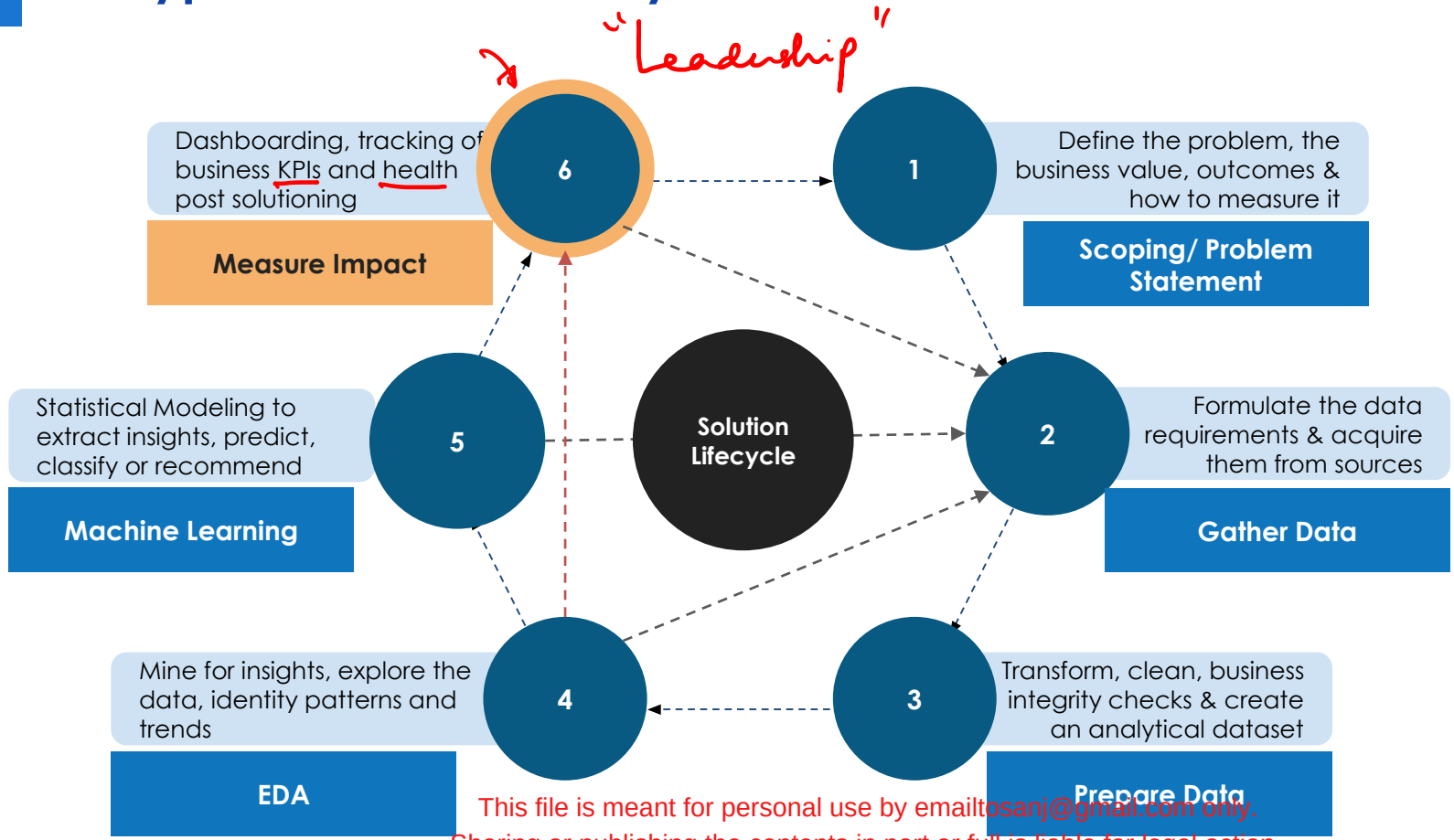


Date	C-ID	Category	Query	Response	Satisfied



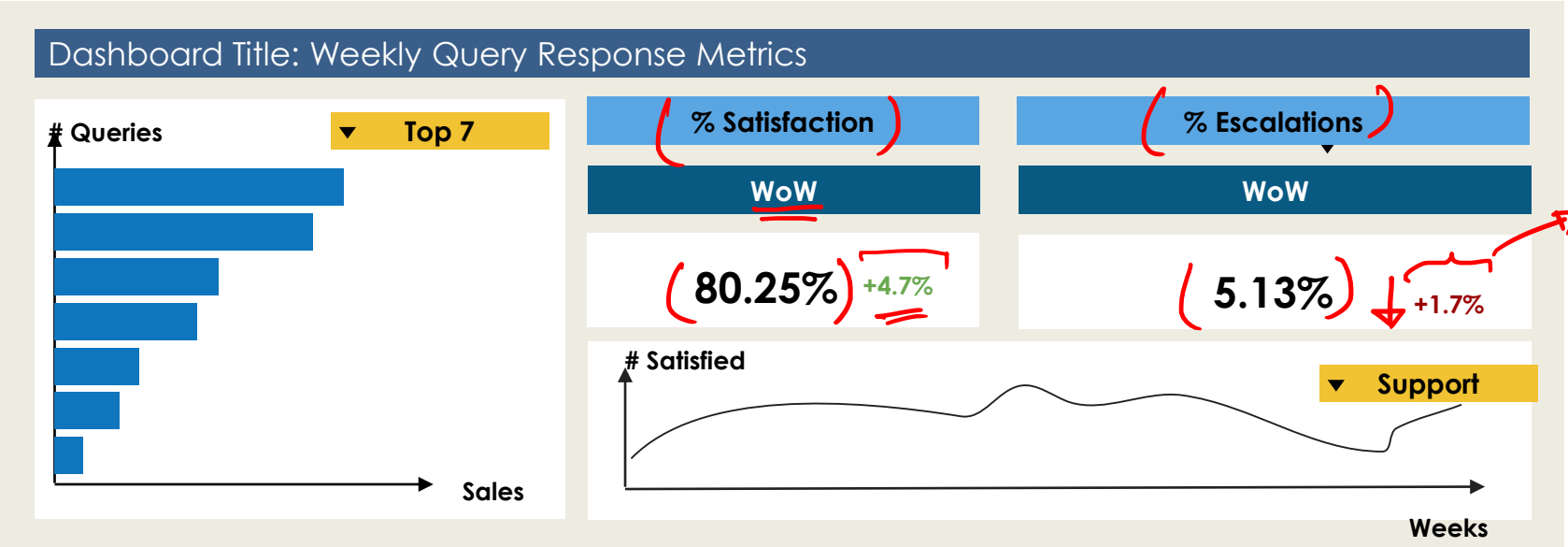
This file is meant for personal use by emailtosanj@gmail.com only.
Sharing or publishing the contents in part or full is liable for legal action.

A Typical Solution Lifecycle



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

Measuring Impact



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

Summary

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 lifecyle 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

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

Great Learning

Happy Learning !

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