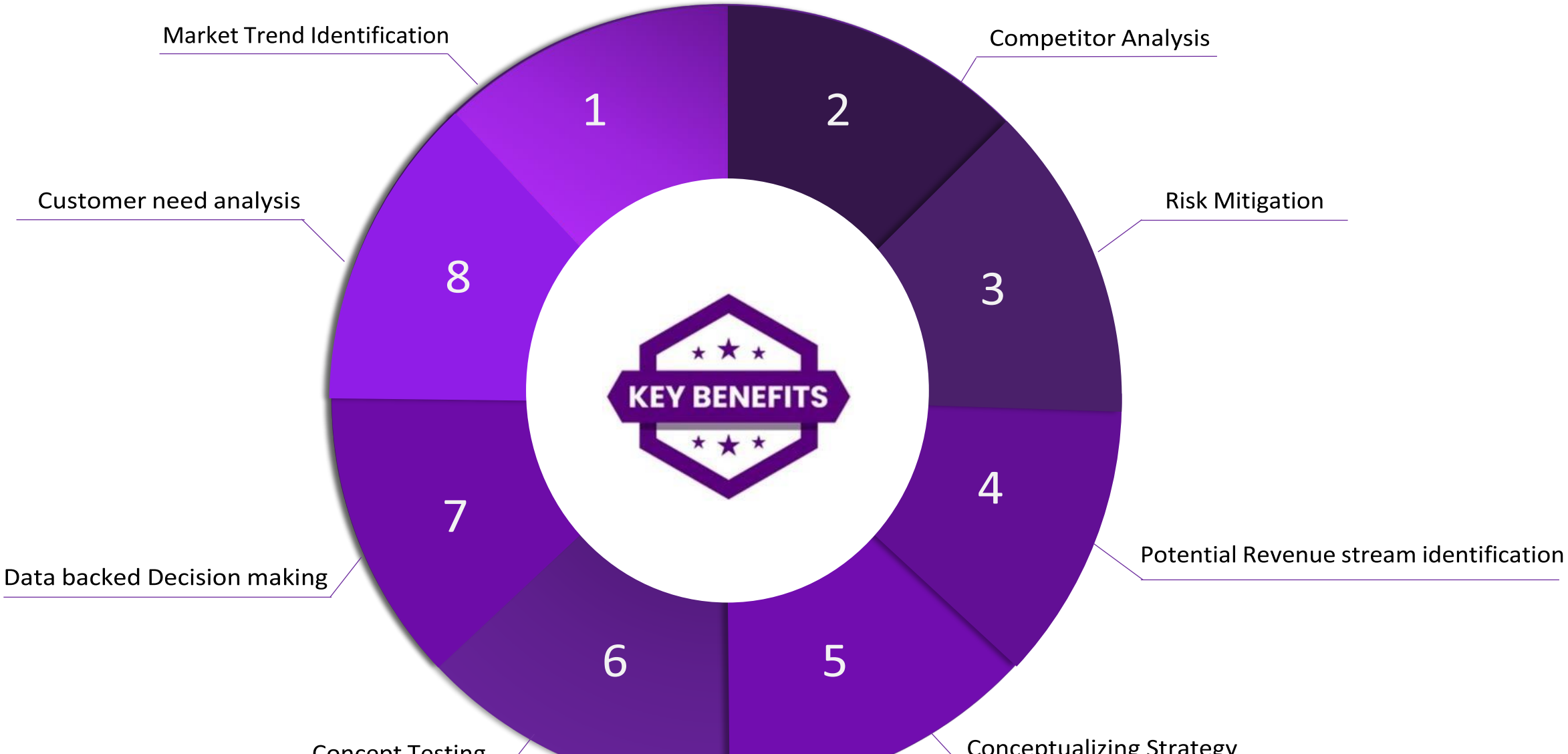
Reimagine Survey Analysis with AI

For Open ended Survey Responses

Table of Contents

What is Market Research

How does Market Research help?



Types of Market Research & Steps

Types of Market Research

Steps Involved

Quantitative Research

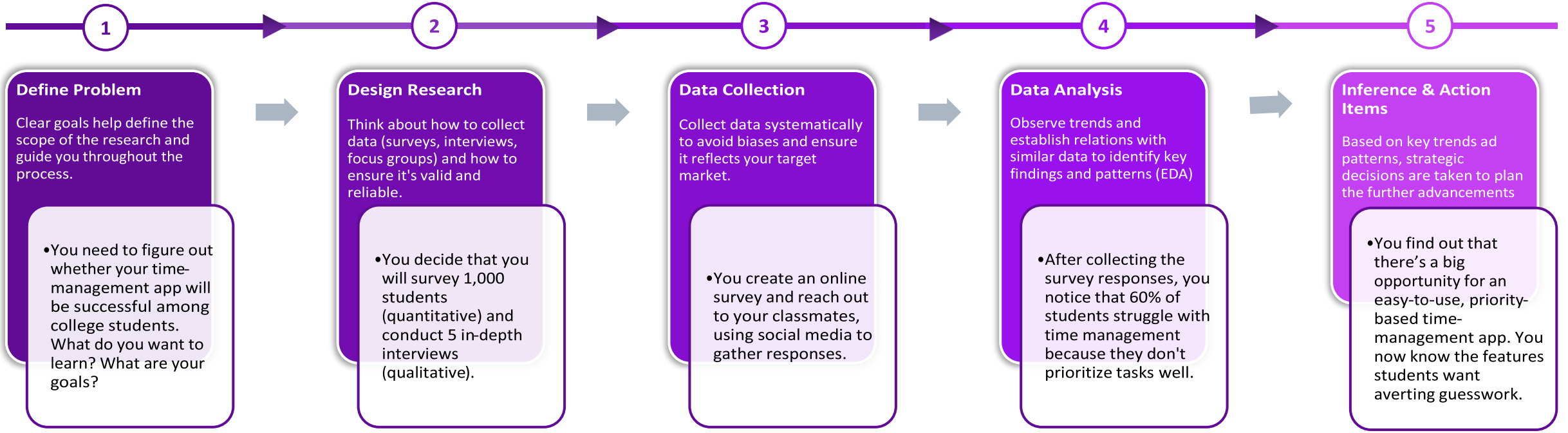
This data gives you numbers that can be analyzed statistically to predict trends. Helps you understand patterns, demand, and general preferences on a large scale.

For Eg: Let's say you conduct a survey of 1,000 students and ask them how often they use a time-management app.

Qualitative Research

This data informs about deep insights into customer motivations, needs, and problems and helps gauge the depth and intensity of the problem or aspect

For Eg: Instead of just asking for numbers, you sit down with a few students and ask them about their biggest challenges with managing time. You learn that some struggle with procrastination, while others feel overwhelmed by their schedule.


1
Define Problem
Clear goals help define the scope of the research and guide you throughout the process.
•You need to figure out whether your time-management app will be successful among college students. What do you want to learn? What are your goals?
2
Design Research
Think about how to collect data (surveys, interviews, focus groups) and how to ensure it's valid and reliable.
•You decide that you will survey 1,000 students (quantitative) and conduct 5 in-depth interviews (qualitative).
3
Data Collection
Collect data systematically to avoid biases and ensure it reflects your target market.
•You create an online survey and reach out to your classmates, using social media to gather responses.
4
Data Analysis
Observe trends and establish relations with similar data to identify key findings and patterns (EDA)
•After collecting the survey responses, you notice that 60% of students struggle with time management because they don't prioritize tasks well.
5
Inference & Action Items
Based on key trends ad patterns, strategic decisions are taken to plan the further advancements
•You find out that there's a big opportunity for an easy-to-use, priority-based time-management app. You now know the features students want averting guesswork.

A Typical Example of Market Research use case Click to learn more

*Imagine you're launching a brand-new app with the goal **to help college students manage their time**.*
You're excited, but you don't know if anyone will use it. How do you find out if this idea is worth pursuing?

To ensure **product's success**, you need to find out the following before even building the app to prevent irrational decisions and potential revenue loss

Definition

Market research is the process of collecting, analyzing, and interpreting **data** about a market, including information about the target audience, competitors, and the overall industry.

Product Demand

Competition

Market Size

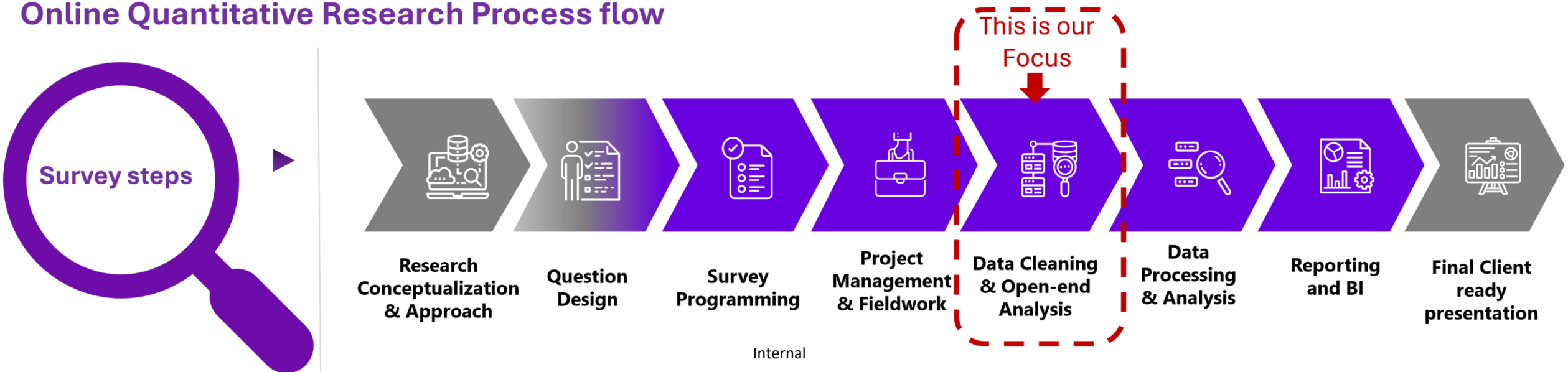
Target Audience

This is where **market research** comes in. It gives you the insights needed to make informed business decisions.

Market Research Methods

setting.

Online Quantitative Research Process flow

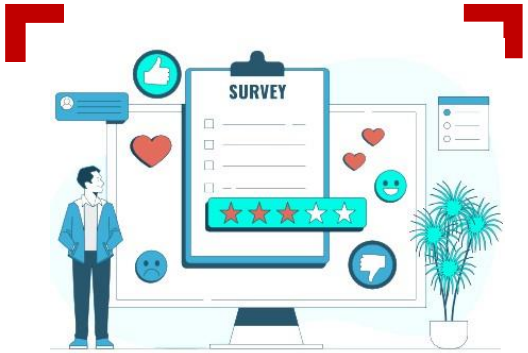


Social Media Monitoring

Analyzing what people are saying about relevant topics

Focus Groups

A small group of people discussing the product or



Surveys

Online or paper-based, used to gather large amounts of online. service in a guided data quickly.

Click to learn more

Observational Research

Watching how consumers interact with products in realworld settings.

Interviews

One-on-one or small group discussions to explore deeper insights.

Why survey quality matters in survey data?

Questionable Market Research Data Quality

- To collect responses in quantitative research, there are panel providers who have database of people who are recruited & consented to take surveys. Any research project reach out to these panel providers to reach out to their panelist as per certain target audience criteria's. The panelist get rewarded with points for taking every successful surveys.

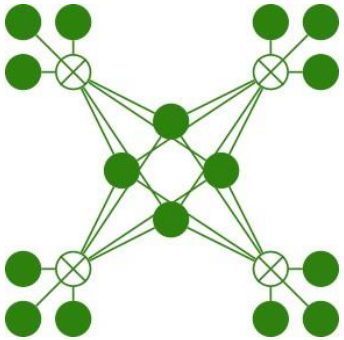
- The entire market research agency is grappling with increasing amounts of panel data that appears to be generated by bad actors – fraudulent survey responses may be originated by (organized) individuals that aim to benefit from accumulated panelist rewards or – this is the much bigger concern – by automated tools and bots that infiltrate research panels.

- Easily available AI applications, such as ChatGPT, Gemini, Copilot, etc., allow non-qualified respondents to come-up with sensible looking responses, even to open ended questions, making it even harder to differentiate legit responses from 'bad' responses.

- Awareness of this problem across the industry is elevated to a level that exposes even perfectly legit respondent data to the suspicion of fraud when results don't match expectations – essentially defeating the purpose of market research.

- As a result massive amounts of manpower is spent on the detection of suspicious patterns in survey data – the process is generally extremely time consuming and often led to inability to identify bot responses.

Impact of Bad Quality Actors on Market Research



Skewed data

Irrelevant Responses

1

Inaccurate conclusions

2

Reduced Credibility

Increased data cleaning and

3

Increased costs

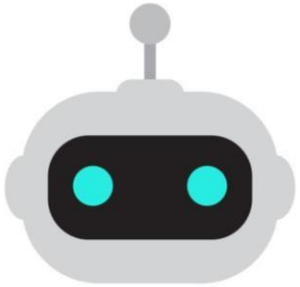
validation.

Different Sources of Survey Quality Flags

To collect responses in quantitative research, there are panel providers who have database of people who are recruited & consented to take surveys. Any research project reach out to these panel providers to reach out to their

panelist as per certain target audience criteria’s. The panelist get rewarded with points for taking every successful surveys.

Common sources



Bots

Scripts programmed to fill surveys automatically, often used to exploit rewards like responses for monetary gift cards.

Click to learn more

Survey Farms

Human-operated setups that mass-submit survey responses for monetary gain.

Click to learn more

Inattentive responders

Genuine users who randomly select answers or fail to read questions thoroughly.

Click to learn more

Open-Ended Spam Reponses & Examples

Quantitative surveys often include open-end questions where respondents provide text responses. These responses serve as valuable data points but can also help identify non-genuine respondents who might compromise data quality.

Filtering out low-quality comments ensures that our data analysis is based on meaningful and relevant responses, leading to more accurate and actionable insights for our clients.

Types

Nonsensical Text

Random strings of words or characters with no coherence.

Example

Question: *What improvements would you suggest for our product?*

Response: *asdfghjkl* or *Lorem ipsum dolor sit amet.*

Copy-Pasted Content

Responses copied from unrelated online sources or repeated text from other survey fields.

Example

Question: *Describe your experience with the service.*

Response: *This is a placeholder text used in web design.*

Repetitive Phrases

Identical responses submitted across multiple open-ended questions or repeated within the same survey across multiple OE questions

Example

Question: *What did you like about the product?*

Respondent 1: *The product provided fulfilled my needs*

Respondent 25: *The product provided fulfilled my needs*

Superficial Responses

Answers that are too brief or vague to provide any meaningful insight. Or is irrelevant to the question asked

Example

Question: *What could we improve about our service?*

Response: *It's fine.* or *Good.*

Responses in a language different from the survey's primary language, often nonsensical or irrelevant.

Example

Question: *What additional features would you like to see?*

Response: 很好. (*Chinese for 'Very good,' in an English survey.*)

Hackathon Details

The Challenge: Elevating Survey Quality

Basic Quality Issues

Detect and filter out responses with:

- Gibberish text
- Profanity or inappropriate content
- Out-of-topic responses

Contextual Quality Issues

Analyze responses for:

- Relevance to question
- Alignment with claimed expertise
- Consistency with facts/knowledge

Bot/AI Generated responses

Leverage advance models to:

- Detect AI-generated responses
- Understand semantic relevance
- Analyze contextual embeddings

Hackathon Structure: From Concept to Execution

Phase 1: Concept & Strategy

Teams brainstorm solutions and develop project proposals.

Phase 2: Presentation

Teams present solutions to evaluation team.

Phase 3: 2-Day Hackathon

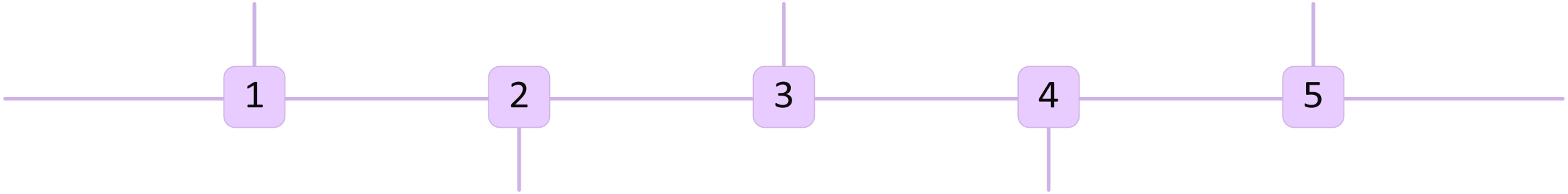
Teams develop AI solutions on real life data.

Phase 4: Evaluation

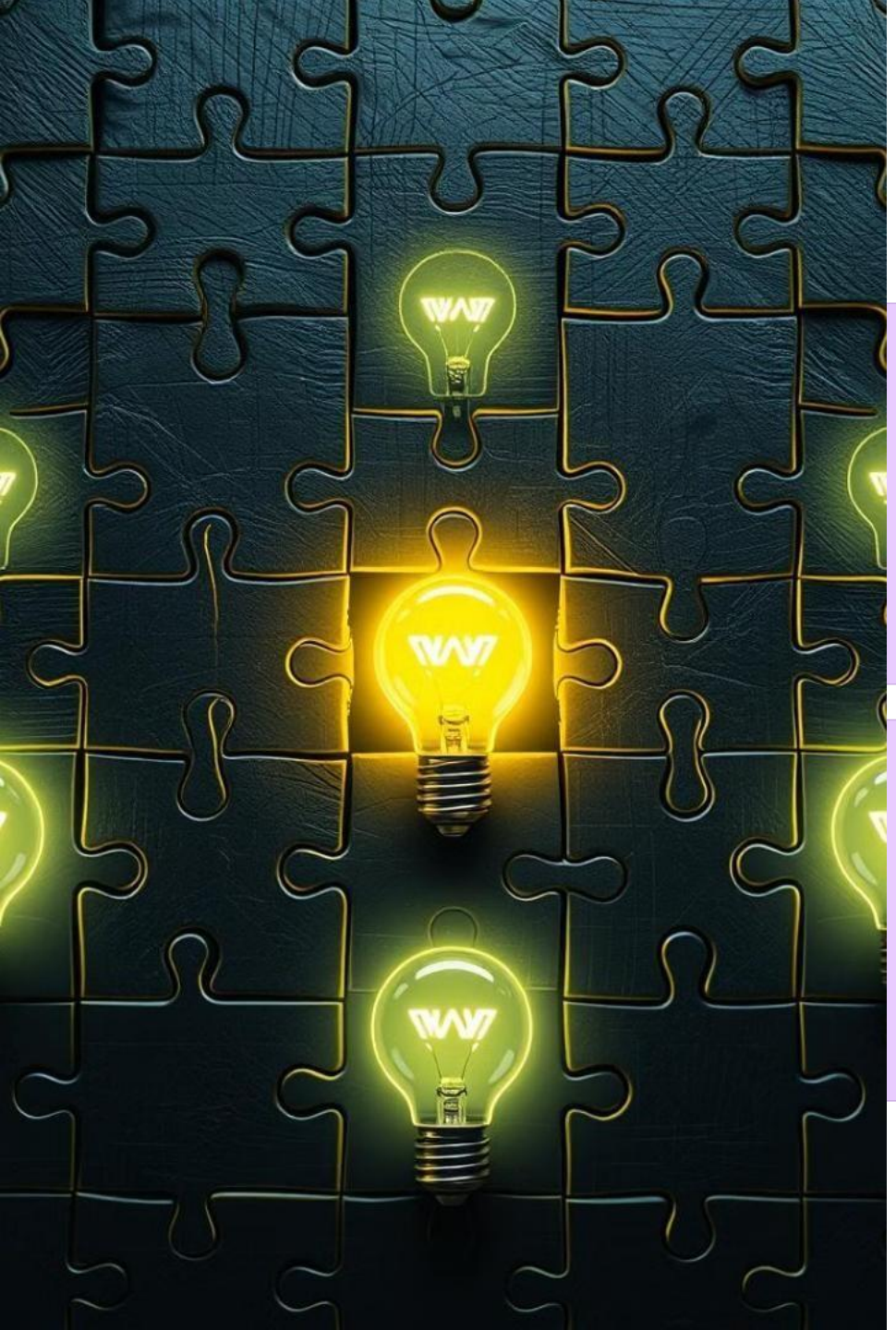
Evaluation of code quality and solution to real-world scalable solution

Phase 5: Implementation

Final team selection to implement PoC



What We're Looking For: Exceptional Solutions



Problem Understanding

Demonstrate a deep understanding of the challenges in validating open-ended survey responses.

Vision and Goals

Articulate a clear vision for the solution and outline specific goals.

Solution Approach

Present a well-defined solution approach that leverages appropriate AI/ML and LLM technologies.

Differentiation

Highlight what makes your solution unique and valuable compared to existing approaches. Innovation in solution would be a key factor

Problem Statement & Data Description

Problem Statement

Develop an AI-powered system that automatically validates and assesses the quality of open-ended text responses in market research surveys. The system should be able to identify and flag low-quality responses including gibberish text, bot-generated content, off-topic answers, copy-pasted content, and inappropriate language. This will replace the current manual review process where human reviewers have to read through thousands of responses to ensure data quality.

Key Themes

Response Quality Analysis: Develop robust mechanisms to detect and filter out poor quality responses including gibberish text, copy-pasted content, and inappropriate language, ensuring data integrity.

Intelligent Context Understanding: Create models that understand the relationship between questions and answers, ensuring responses are relevant and meaningful within the survey context.

AI/Bot Response Detection: Implement sophisticated detection methods to identify artificially generated responses, including those from advanced language models and automated bots.

Automated Validation Interface: Build a real-time system that can process large volumes of responses automatically, with configurable quality thresholds and detailed reporting capabilities.

Survey Structure

A concept testing survey evaluating consumer reactions to ULTRA Pure Gold, a new premium light beer positioned as a 'Superior Light Beer' with low carbs (2.6g) and calories (85), among adults aged 21-65 who consume alcoholic beverages.

Demographics:

- Age
- Gender
- Location (Urban/Semi-urban/rural)
- Income

Category Behavior:

- Current alcohol consumption
- Beer type preferences
- Consumption patterns

Category Behavior:

- Product shown to respondents and asked a series of questions
- Relevance
- Appeal
- Differentiation
- Believability
- Price perception
- Purchase intent
- Expected consumption

Data Overview

The dataset contains survey responses evaluating a new ULTRA Pure Gold beer concept. The data is organized across three Excel sheets, each providing different views of the same information:

1. **Data Set with Labels Text** - Contains responses with full text labels
2. **Data Set with Values as answer** - Contains the same data with numeric codes
3. **Data Dictionary** - Detailed variable descriptions and answer options

Key Variables for Model Building **Target**

Variable:

OE_Quality_Flag (Binary: 1 = Flagged for Quality, 0 = Not flagged for Quality)

- Present in training data
- To be predicted in test data

Primary Input Variables (Open-End Responses):

- **Q16A** - Likes about the concept (What respondents LIKE about the shown product)
- **Q16B** - Dislikes about the concept (What respondents DISLIKE about the shown product)
- **Q18_1** - First product replacement choice
- **Q18_2** – Second product replacement choice
- **Q18_3** – Third product replacement choice

Data Format

- Each row represents one respondent
- Variables are mix of categorical, ordinal, and open-text
- Refer to Data Dictionary sheet for complete variable descriptions and answer options
- Numeric codes in "Values" sheet correspond to text labels in "Labels" sheet

Sample Data

Sample Data for Hackathon.xlsx

Ideation Submission

- A PowerPoint presentation ideally detailing:
- Problem & Data Understanding
- Solution approach
- Proposed model or technologies architecture
- Differentiation/Unique proposition of the proposed solution
- Scalability of the proposed solution
- Key challenges the team is anticipating

Hackathon Final Submission

- No restrictions on programming languages or frameworks
- Solution must be executable on standard computing infrastructure
- Code must be well-documented and include setup instructions
- Final submission must include:
  - Source code
  - Requirements file
  - Setup documentation
  - Brief user guide & demo video

Performance metrics on test data