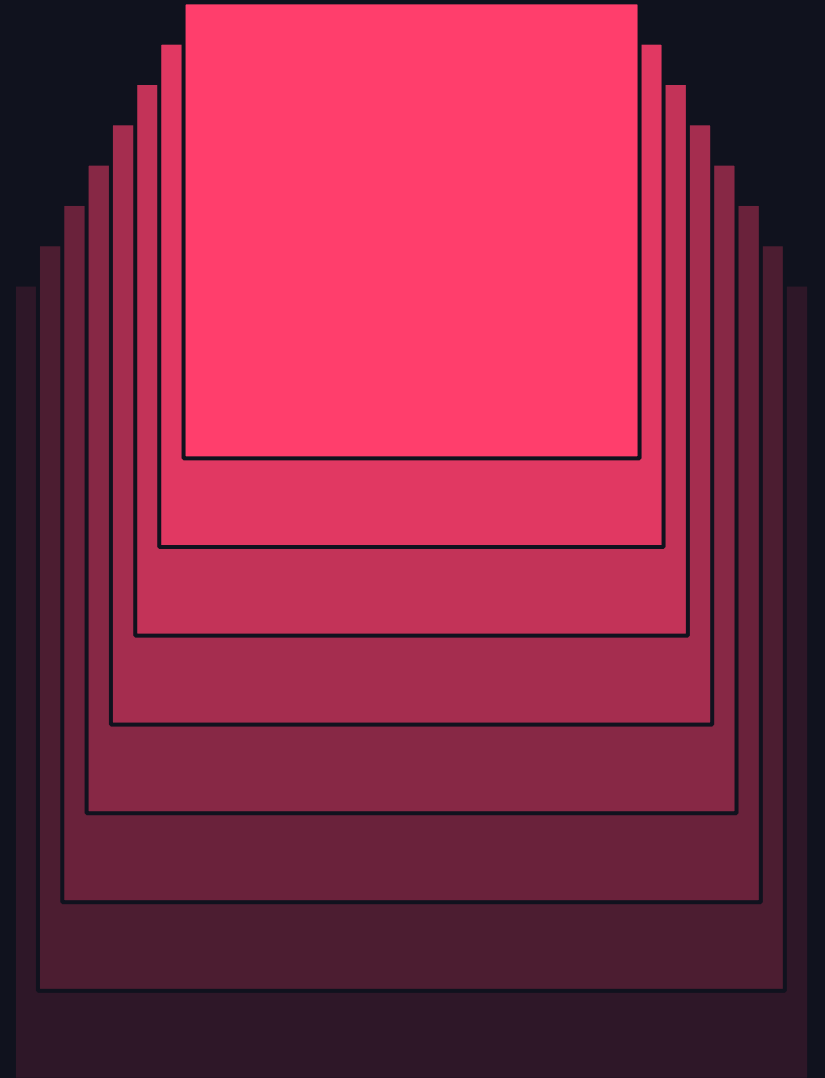# DESTINATION DATA: UNPACKING TRAVELER INSIGHTS WITH LLMS

Redkite
June 2024

Introduction

Sonia Patel

**Generative AI Lead
Redkite**

7 years of experience harnessing emerging technologies to deliver value to organisations through data-driven excellence and insightful thought leadership

Gaziz Zhotabayev

**Data Science Lead
Redkite**

Over 8 years of experience in data solutions and leading analytics teams, helping organisations achieve strategic business objectives and realise value through data

REDKITE Part of **Accenture**

The global pandemic disrupted the travel industry

Pandemic changed travel patterns

01

How and why people travelled before the pandemic and how they do so now has changed significantly, which means **companies need to adapt their approach** to reach their target audiences

## Primary research helps answer the 'why'

02

While mobility and other traveler behaviour data are a good proxy that help understand the impact of marketing efforts, it is surveys that offer **crucial insight on traveller sentiment** and the **real "why"** behind their travel decisions

Traditional analysis leaves insight on the table

03

Focussing on pre-coded multiple-choice questions at the expense of open-ended ones **risks ignoring critical bits of insight**

Resulting in a need for a new approach

04

What if we could make free-form text data as analyzable as numerical data?
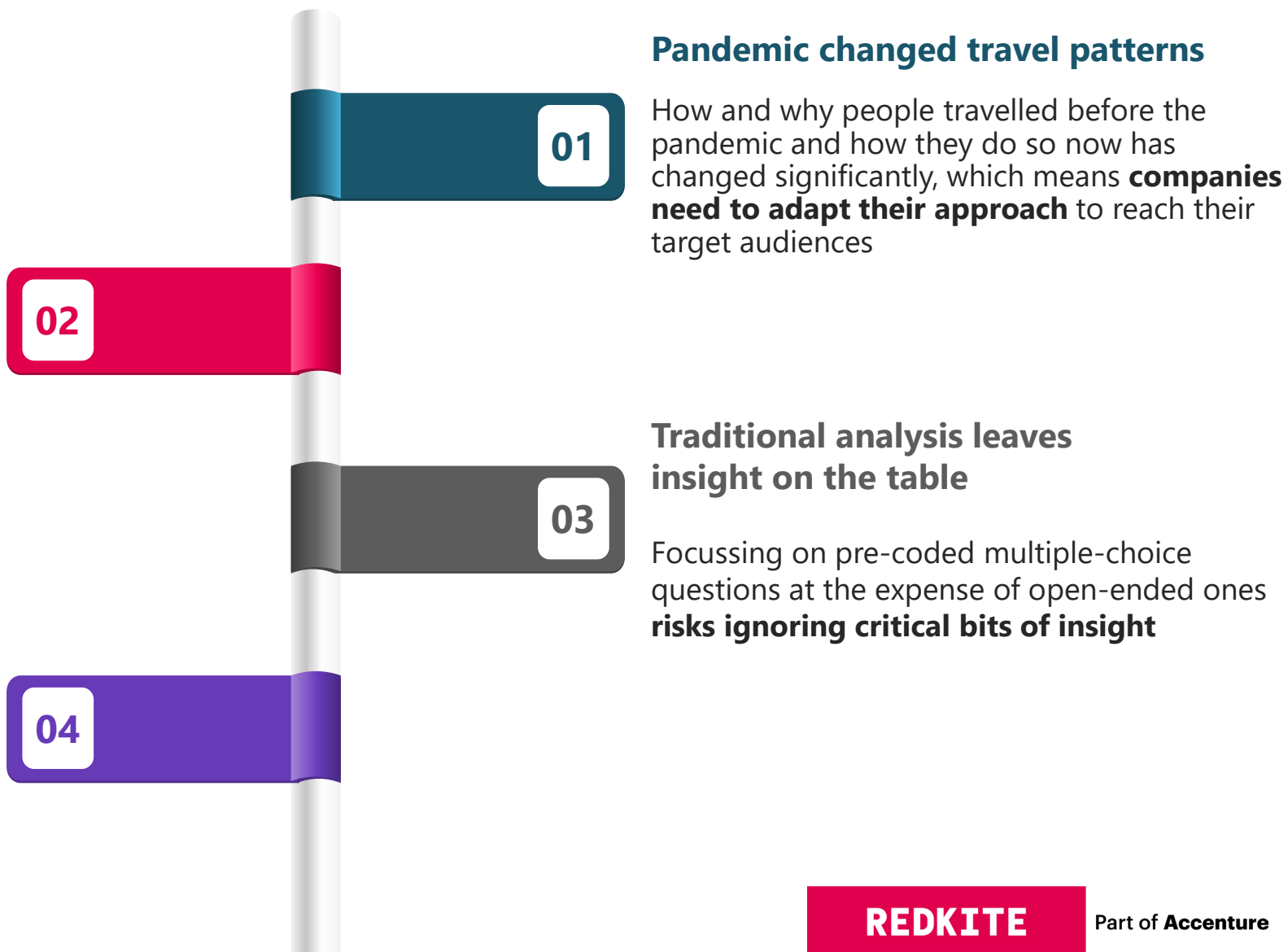
# Delving deeper into the situation our client was experiencing

MISSION
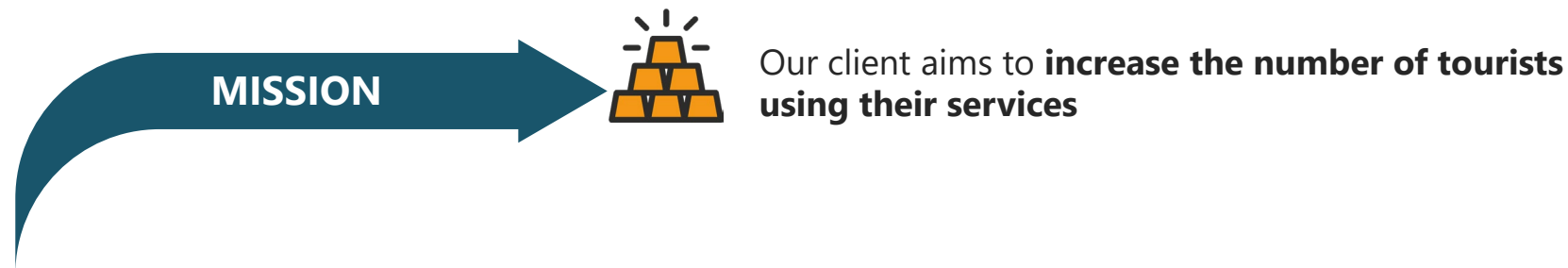
Our client aims to **increase the number of tourists using their services**

ENABLER

They use advertising campaigns to engage and inspire the audience to visit. To evaluate campaign impact and support with campaign planning they, among other things, conduct market research that **generates vast volumes of free-text responses from tourists**

BARRIER

Reviewing free-text responses is an **incredibly manual task**, meaning **analysis can be inconsistent** and **time-to-insight is often too long to affect change**

DESIRED OUTCOME

Solution needed to **streamline the labor-intensive** analysis processes, **improve analytics**, and, as a result, **bolster campaign-related decision-making**

REDKITE Part of **Accenture**
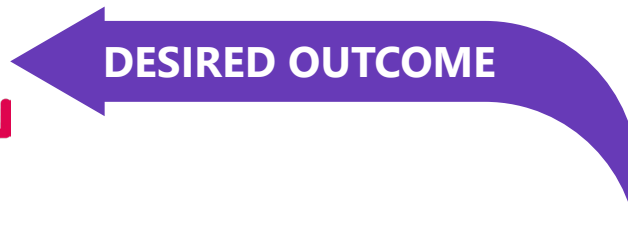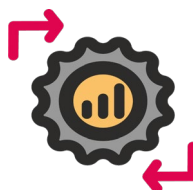
Now let's look at the challenges our client faced with survey responses

Untapped insights

Difficulty extracting insights from open-ended responses results in a **struggle to gain nuanced perspectives and sentiments from tourists**

Labor-intensive process

Reviewing free-text responses can take **days of effort**

Unreliable pre-coded responses

Pre-coded survey answer options can result in **biased and inaccurate responses** and a longer survey-iteration cycle

Non-informative responses

A lot of responses are **non-informative** and need to be **filtered out**

Multi-language complexity

Multiple languages adds an **extra layer of complexity** as each review needs to go through a translation process

Emerging data capability

To analyze free-text responses we need not only **appropriate tooling**, but a data capability **suited to and right-sized** for the business

REDKITE Part of **Accenture**

Here are some example of free-text responses received from a survey

| Example survey questions	Example answers
Please tell us, in your own words, why you are unlikely to consider [destination] for a holiday or leisure break?	"Because it is difficult to obtain a visa"
List the first words that come to mind when thinking about a holiday or short leisure trip.	▪ Sun ▪ Europe ▪ Price ▪ Cuisine
In the context of a holiday or short leisure trip what does the word 'Adventure' mean to you?	"Adventure means to me that you discover different things and experiences you encounter in different places. Learn new things and have fun"
Please now describe, overall, how the campaign made you feel about [destination] **(Our focus for this talk)**	"I think it's great when you can do a lot more in a country than check out the sights. This seems much more personal, the ad takes the cliches and reinterprets them in a funny way."

LLMs or traditional NLP methods?

We chose LLMs for most tasks over traditional NLP methods for **four key reasons**

Time-to-deploy
LLMs are more likely to work out-of-the-box and are not as reliant on specialist skillsets
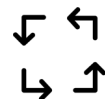
Reusability
LLMs can naturally extend to other use cases that involve text analytics

Summarization
LLMs are better at accurately summarising the intended meaning of the input text
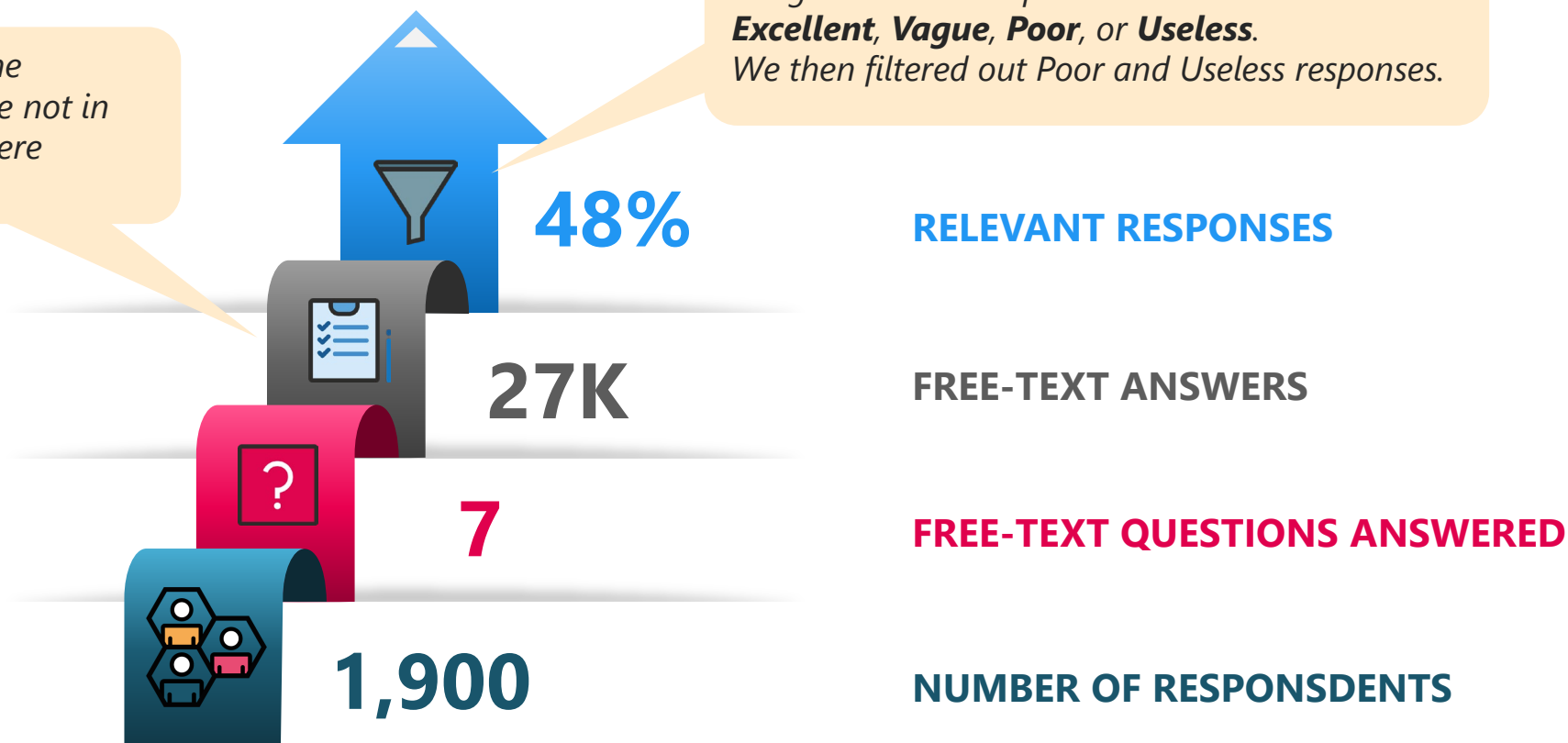
Generalizability
Traditional NLP methods often require re-training whilst LLMs need good prompting

First, we prepped the data by translating it and filtering out low-quality responses

Example Survey Response: *"mir fällt nichts ein" >> **Translated** >> "I can't think of anything"*
Relevancy Score: *1 (Useless)*

*Over half of the responses were not in English and were **translated***

*We graded each response's relevance as either **Excellent**, **Vague**, **Poor**, or **Useless**. We then filtered out Poor and Useless responses.*

48% **RELEVANT RESPONSES**

27K **FREE-TEXT ANSWERS**

7 **FREE-TEXT QUESTIONS ANSWERED**

1,900 **NUMBER OF RESPONSDENTS**

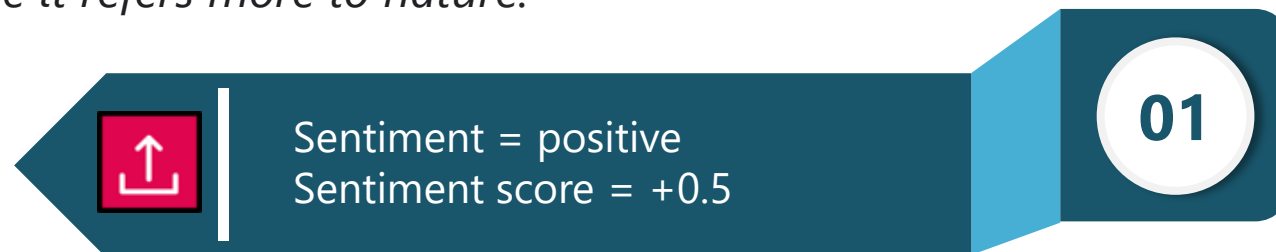*Average figures per single survey wave*

For each response, we looked at the sentiment, emotion, and topics mentioned

Example Survey Response

"A quirky and fun style of advertising that I really like. On the other hand, festivals, graffiti [do] very little for me! the second part of the ad would suit me better since it refers more to nature."

SENTIMENT ANALYSIS
Classify (into positive, negative, conflicted, neutral) and score (with a -1 to 1 score) the sentiment of each response

Sentiment = positive
Sentiment score = +0.5

01

EMOTION CLASSIFICATION
Classify the emotional content of the response

Emotion = joy

02

TOPIC EXTRACTION & CLASSIFICATION
Tag topics mentioned in the response, including a prevalence score, and classify topics into pre-defined categories

Topic = Ad (3), Culture (1), Things to see/do (2)

03

TOPIC-LEVEL SENTIMENT ANALYSIS
Classify the sentiment polarity of each topic extracted from the response

Ad: positive
Culture: conflicted
Things to see/do: positive

04

Some of the critical responses are where the most interesting insights came from

Example Survey Response

"I mean, I didn't need the floating distracting head. [This country] has so much to offer, quirky floating heads are not going to sell it to me, price will always win there."
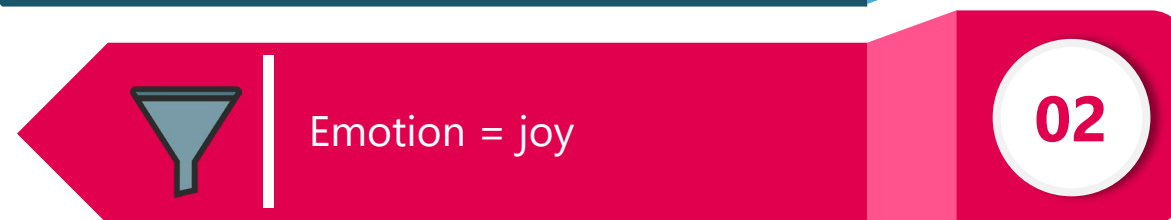
SENTIMENT ANALYSIS
Classify (into positive, negative, conflicted, neutral) and score (with a -1 to 1 score) the sentiment of each response

Sentiment = negative
Sentiment score = -0.38
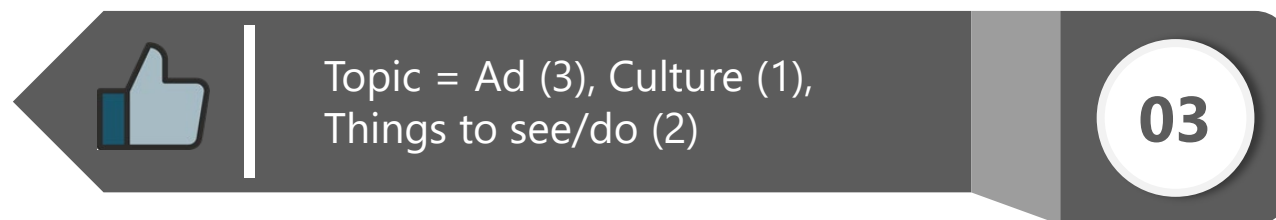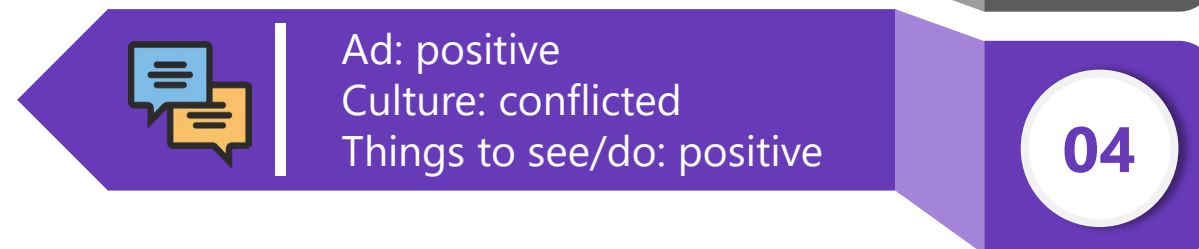
01

EMOTION CLASSIFICATION
Classify the emotional content of the response

Emotion = indifference

02

TOPIC EXTRACTION & CLASSIFICATION
Tag topics mentioned in the response, including a prevalence score, and classify topics into pre-defined categories

Topics = Affordability (3), Ad (3), Things to see/do (3)

03

TOPIC-LEVEL SENTIMENT ANALYSIS
Classify the sentiment polarity of each topic extracted from the response

Affordability = neutral
Ad = negative
Things to see/do = positive

04

REDKITE Part of **Accenture**

Finally, we asked LLMs to create topic-level and campaign-level summaries of responses

Affordability
Holiday experiences
Culture and social perspectives

Response	Topic	Prevalence	Sentiment
A dream of mine to visit the castles	Castles	2	Positive
Love the accent	Accent	3	Positive
That recognisable and funny humour	Humour	3	Positive

Topic-level campaign summary	
Topic	**Summary**
Culture and social perspectives	[The country's] culture, humour, and accents received overwhelmingly positive feedback, with humour being particularly appreciated.
Holiday experiences	Respondents reminisced fondly about their stays, citing beautiful landscapes, enjoyable moments, and a friendly welcome from locals.
Affordability	Overall, the survey reflects a concern over the costliness of certain experiences, while recognizing some positive aspects of value and accessibility.

Campaign-level summary of summaries

Holiday experiences in [the country] were fondly remembered for their fun and beauty, though a few had negative impressions... Cultural and social perspectives were mostly positive, celebrating [the country's] culture and humour... Affordability was met with mixed feelings; positive views on value and accessibility were tempered by concerns over expenses... Overall...

How do we ensure our solutions are good enough?
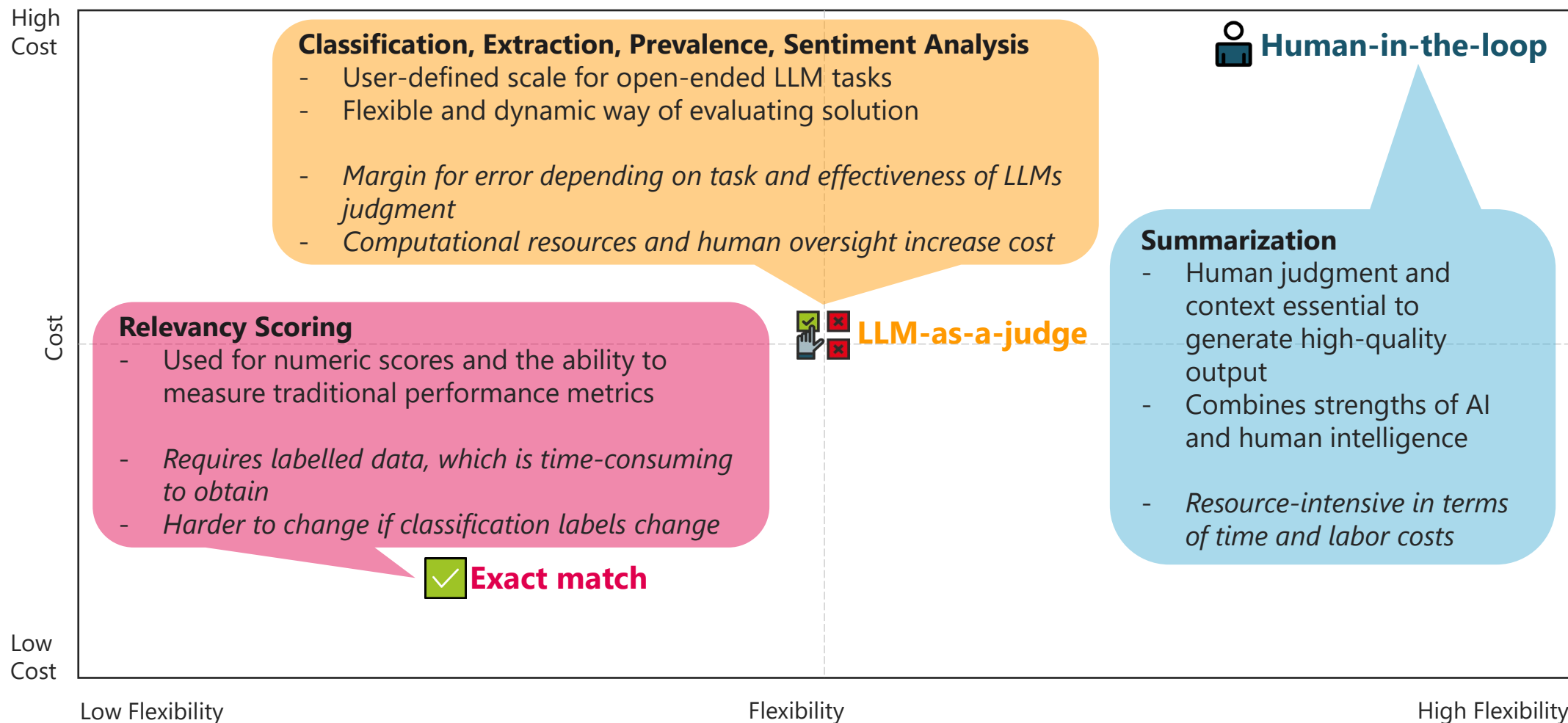
Databricks' flexible approach enabled a successful build of the solution

Individual Response

Unity Catalog

Feature Table

Question	Answer
How do you feel about [the country]	Great, I really liked the museums and food

Chosen Model
GPT-4

"You are a data specialist, you are tasked with..."

Model Execution

Derived Table: Response Topic Level

Topic
Museum
Food
...

Logging Table

LLM Metric	Model	Input Tokens	Output Tokens	Latency	LLM Output
Topic	gpt4-topic-v5	400	4	0,5	Museum;Food
...	...	...	...	...	...

1.

2.

Future Update
External endpoint development was made available in April 2024 which enables:

1. Allows us to switch between models more easily in the future

2. Seperate logging table replaced by inference tables in Databricks

Digging deeper into the value Databricks brings to this solution

(2) Experiment view (built on top of Mlflow) made it easy to compare results and drill into side-by-side comparison when evaluating model building
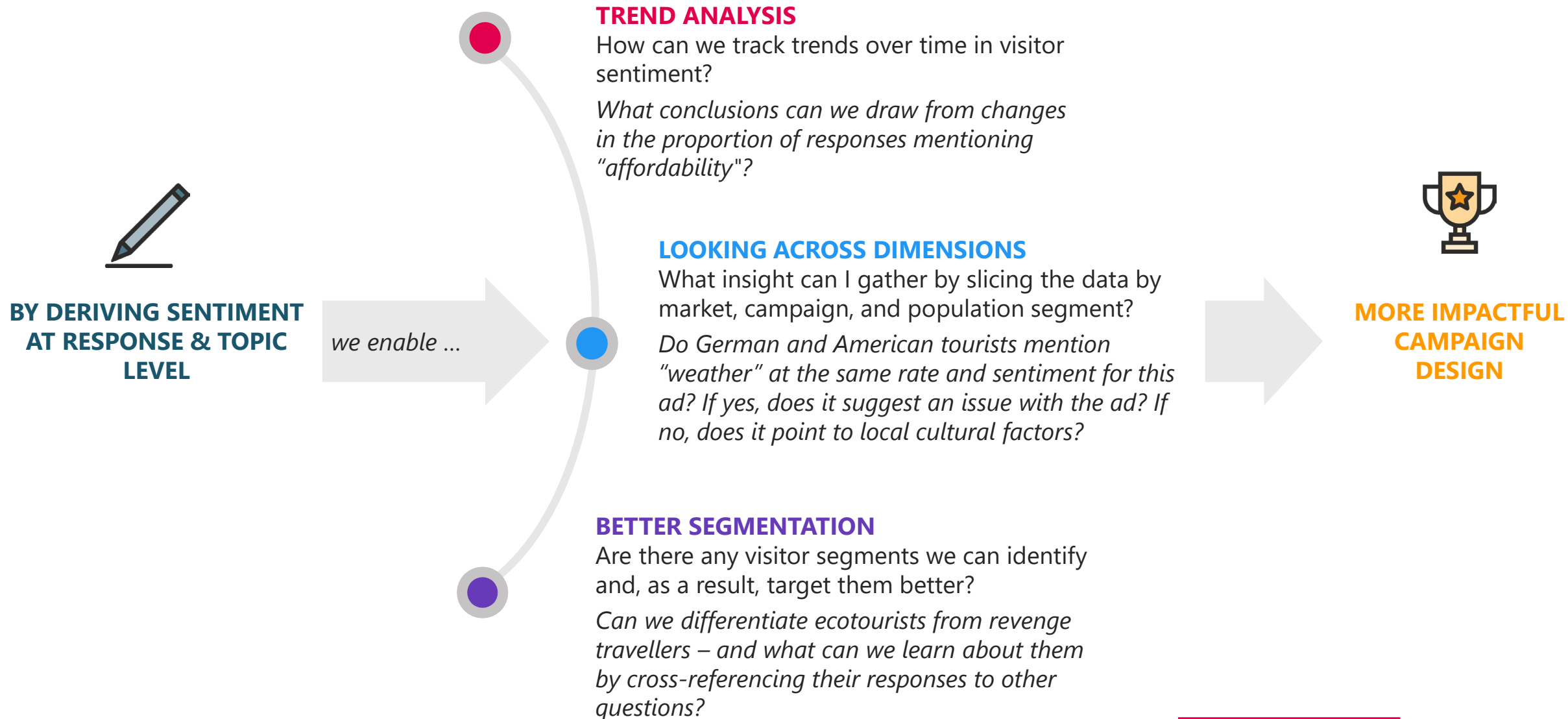
(1) Unity Catalog streamlines storage and segmentation of data, models, and outputs, enabling permission-based access and seamless querying of PROD data from DEV environments

ml*flow*
Evaluation Methods

(3) LLM-as-a-judge allows a low code way to evaluate data at scale

LLM Models

ml*flow*
Prompt Templates

ml*flow*
Log outputs

ml*flow*
Evaluate

Feature table

LLM Transformation

ml*flow*
Output dataframe

Output table
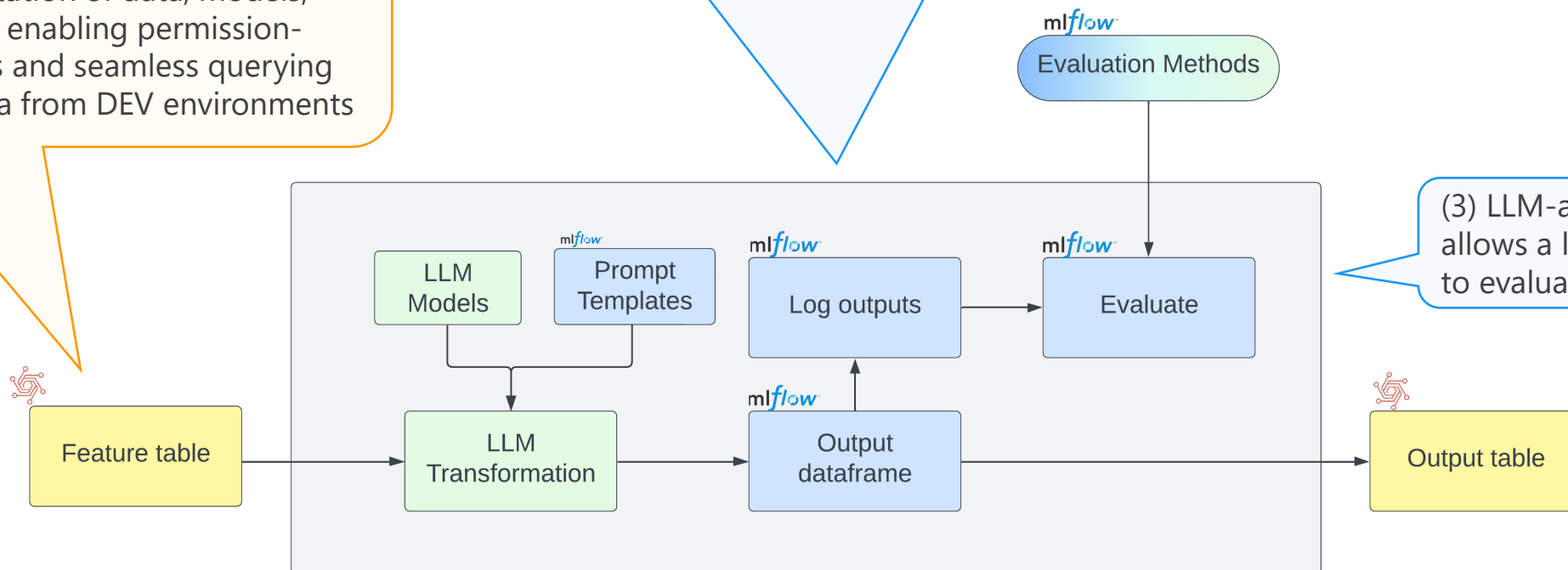
(4) External endpoint for future iterations
This allows for easy switching between models, creates an inference 'logging table' automatically and allows prompt engineering to be carried out from UI vs needing notebook, lowering the barrier needed to build code

Deriving response-level fields opens up a wealth of analytics possibilities in the aggregate

BY DERIVING SENTIMENT AT RESPONSE & TOPIC LEVEL

we enable ...

TREND ANALYSIS

How can we track trends over time in visitor sentiment?

What conclusions can we draw from changes in the proportion of responses mentioning "affordability"?

LOOKING ACROSS DIMENSIONS

What insight can I gather by slicing the data by market, campaign, and population segment?

Do German and American tourists mention "weather" at the same rate and sentiment for this ad? If yes, does it suggest an issue with the ad? If no, does it point to local cultural factors?

BETTER SEGMENTATION

Are there any visitor segments we can identify and, as a result, target them better?

Can we differentiate ecotourists from revenge travellers – and what can we learn about them by cross-referencing their responses to other questions?

MORE IMPACTFUL CAMPAIGN DESIGN

What value does this solution bring?

TIME BACK TO BUSINESS

Transforming a historically slow and arduous task into a process that handles large volumes of survey responses quickly and efficiently gives employees time back to focus on more value-add activities
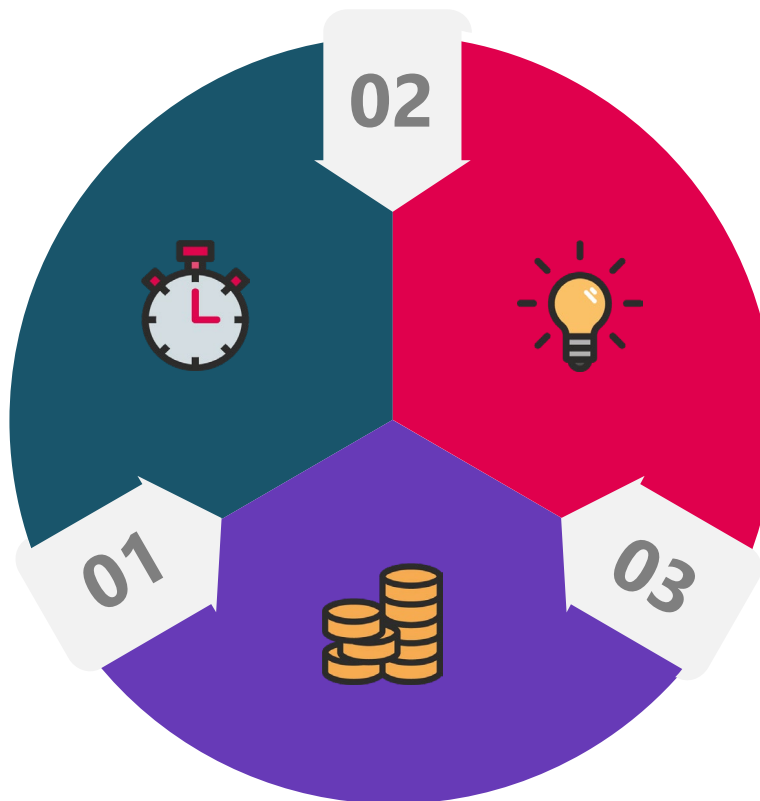
RICHER AND TIMELIER INSIGHTS

LLMs help derive standard and scalable data points such as sentiment, emotion, topics mentioned, and others, helping uncover patterns within the data that support decision-making

HIGHER ROI ON CAMPAIGNS

Ultimately, richer insights allow for better-targeted strategies, leading to more impactful and effective campaigns

REDKITE Part of **Accenture**

16

How will LLMs impact the future of surveys?
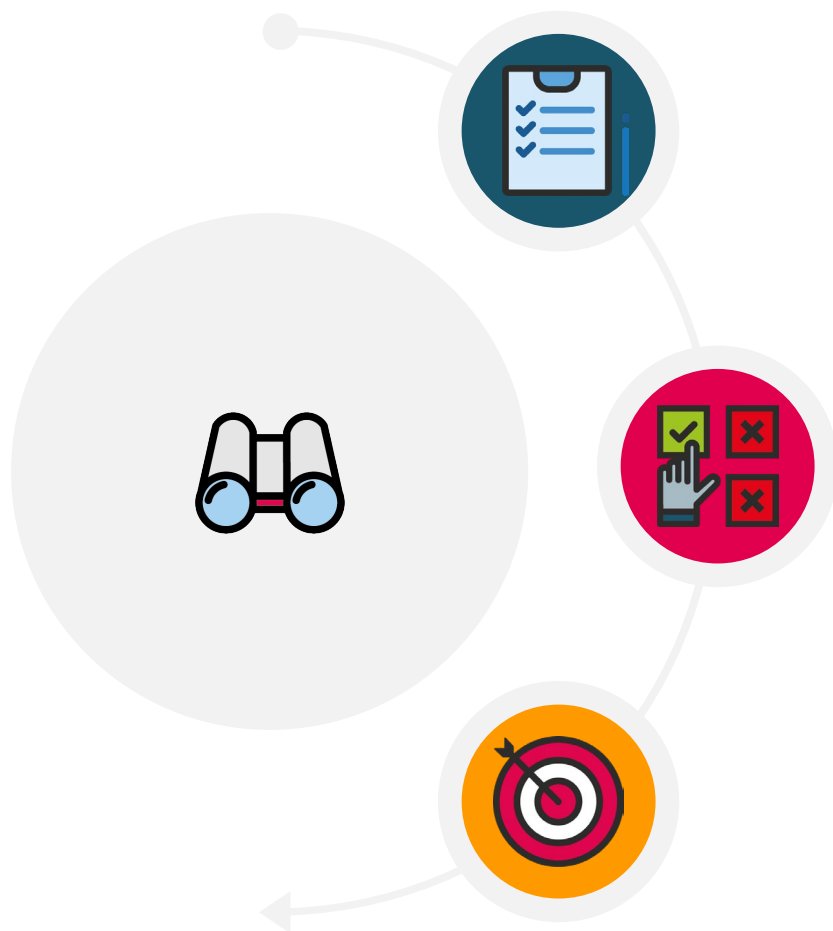
BROADER APPEAL AND DEEPER INSIGHTS
Making direct feedback and qualitative insights readily available for analysis, leveraging free-text responses will become a standard tool in Marketing's toolkit

SIMPLER AND FASTER SURVEY DESIGN
Using LLMs to reduce the reliance on getting pre-coded options 'just right' during design combined with developing multimodal capabilities (e.g., speech-to-text) will enable quicker survey iterations, allowing marketing strategies to remain relevant and responsive

A 360-VIEW OF YOUR TARGET AUDIENCE
Augmenting quantitative data with qualitative insights derived from surveys will allow organizations to get a deeper understanding of their customers, providing a richer basis for developing strategies to reach them

Reach out to us!

CONTACT REDKITE

To hear more about our generative AI solutions and how we can work together help you and / or your clients

Sonia.Patel@redkite.com

Gaziz.Zhotabayev@redkite.com

SCAN FOR THOUGHT LEADERSHIP

Scan the QR code to find some compelling thought leadership pieces for your industry

DATA⁺AI SUMMIT