talkscape.

Technical Paper:
The Talkscape Data Journey

www.talkscape.com

Introduction

The Talkscape platform has been designed with data security, client sovereignty, and privacy at its core. This document outlines the full data journey. It begins with the creation of secure, client-specific AWS environments, through to the processing, analysis, and presentation of feedback insights. It details how voice data is captured, transcribed, analysed, and transformed into actionable insights, while ensuring that all processing steps are encrypted, controlled, and compliant with leading data protection standards. By using a tenant-based architecture, transient external API processing, and strict non-retention configurations, Talkscape guarantees that each client's data remains fully segregated, private, and under their control throughout the journey.

Setting up a secure data environment

The Talkscape data journey begins with the setup of a secure, isolated AWS environment dedicated to each client. This tenanted architecture ensures that all client data (surveys, spoken responses, transcriptions, and audio files) is stored exclusively in that client's own encrypted S3 buckets. Access to this environment is tightly controlled and can only occur through the Talkscape platform using multi-factor authentication (2FA), adding an additional layer of identity verification. The underlying infrastructure benefits from the robust security controls provided by AWS, including AES-256 encryption at rest, TLS 1.2+ encryption in transit, fine-grained access controls, and comprehensive audit logging. With AWS's compliance with global standards such as ISO 27001, SOC 2, and GDPR, clients can be confident their data is protected according to industry best practices from the outset.

Data capture and preprocessing

Once data is captured through the Talkscape platform, it undergoes two key processing steps: automatic speech recognition (ASR) and optional voice modulation. In both cases, audio files are sent securely using encrypted API calls, and the providers are configured to operate in non-retention mode, meaning they do not store or log the data after processing. The processed outputs are immediately returned to Talkscape and stored exclusively in the client's private, encrypted AWS S3 bucket. No copies of the audio or text are stored on any third-party systems. This ensures that all data processing is transient, tightly controlled, and remains entirely within the client's secure AWS tenant environment.

Theme Generation

We use our proprietary Natural Language Processing model to extract the themes present in the data. From this initial list of themes and sub-themes, the client can refine the model further. This is by relabelling themes, merging themes, or deleting themes completely. They can also search the data for new themes and add them to the model. The custom NLP learns from these changes, with any new data coded against this customised thematic structure.

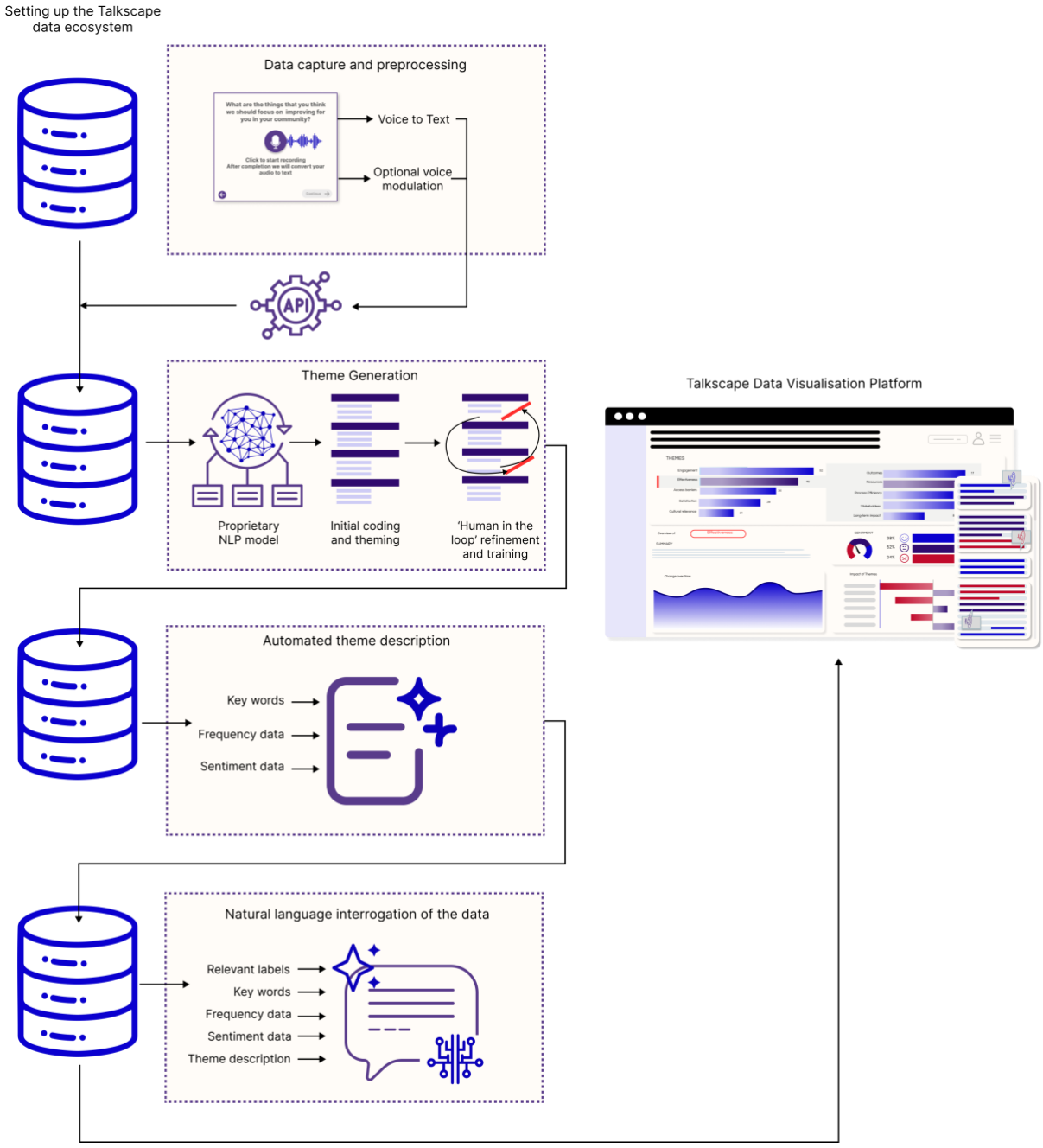
Automated description generation

Once the themes have been formed and labelled, Talkscape produces a natural language summary of the theme. To do this, we use: 1. The frequency of the theme; 2. The sentiment score from each text snippet that relates to the theme; 3. Representative key words or phrases. At no point are the raw quotes for data mapped against each theme, ever leaving the secure environment. The summary data is then processed using a series of prompts to LLMs. We use a combination of AWS Bedrock and Microsoft Cognitive Services for this, as they allow LLMs to be used in a secure way. The data is not used to train their LLMs.

Natural language interrogation of the data

Talkscape allows users to ask natural language questions of the data. For example, a user might type in a question like “What are the most important issues for region 1 versus region 2?”. Where preparing natural language responses to these questions, Talkscape will submit the following data via a secure API to the LLMs we use (AWS Bedrock and Microsoft Cognitive Services). No raw quotes are ever used. What is submitted are the theme descriptions, any relevant labelling data (e.g. in this case, indication of region 1 and region 2), illustrative key words/phrases, and sentiment score range and frequency. From this data, the LLM will produce a natural language answer to the question submitted.

Figure 1: An illustration of the data journey in Talkscape



Conclusion

Talkscape provides clients with a secure, scalable, and transparent method for capturing and making sense of authentic voice data. At every stage Talkscape applies privacy-preserving design principles and ensures that no raw data is exposed or retained by third parties. Only anonymised, structured summaries are ever used to interact with external AI tools, and always in non-training, ephemeral sessions. By combining technical safeguards, privacy-conscious processing, and full client control, Talkscape offers peace of mind for organisations committed to ethical and secure data practices.

Contact Us

Book a Demo

+6427 4191663

hello@talkscape.com