UNIVERSITY OF SURREY

CAQDAS networking project
Computer Assisted Qualitative Data AnalysiS

Department of Sociology

Reveal: Distinguishing features

This document is intended to be read in conjunction with the 'Choosing a CAQDAS Package Working Paper' which provides a more general commentary of common CAQDAS functionality. This document does not provide an exhaustive account of all the features and functions provided by Reveal but is designed to highlight some of its distinguishing elements. The Comment section at the end details our opinions on certain aspects of functionality and usability. See also Silver & Lewins (2014) Using Software in Qualitative Research: A Step-by-Step Guide, Sage Publications and software developer website. This document reviews Reveal as of October 2024. Many thanks to Alok Jain for reviewing this document for accuracy.

Background   www.doreveal.com

Reveal is an online Generative-AI tool designed to facilitate the qualitative research process with a set of transcription and automatic data synthesis tools ■ Developed in the context of supporting market research projects its focus is on participant data gathered via one-to-one interviews and focus-group discussions ■ Reveal is developed by Alok Jain and Niket Patel at Synthefai Inc. in Virginia (USA) and Gujarat (India) and was first available for public use in July 2023

Minimum System Specifications (recommended by developer)

Reveal will run using recent versions of any of the popular internet browsers ■ Currently the application is built for a desktop / laptop / tablet screens, but there are plans to make it available for smartphone screens in the future.

Data security in Reveal   https://doreveal.com/#data-security

Reveal uses the OpenAI API to generate responses ■ Reveal have an agreement with OpenAI such that data uploaded through this API is not used to training OpenAI's models or to improve its service offerings, and that OpenAI will not share uploaded data with third parties ■ Data is encrypted ■ Users of Reveal can request data to be deleted at any time – this includes the research data itself and any prompts etc. used to process the analysis ■ There is an option to store research recordings, transcripts and the AI analysis in EU by using eu.doreveal.com

The Structure of work in Reveal   Analysis happens by creating "studies" within the

platform, and users can create multiple studies ■ The basic unit-of-analysis is a research participant, which is specified after audio/video have been transcribed or prepared transcripts uploaded ■ The platform focuses on the text specified as belonging to participants in generating responses, while using moderator/interviewer words as context ■ Participants can be assigned 'tags' (e.g. to represent socio-demographic characteristics') for filtering and comparison purposes

## Data types and formats   Audio and video files in common formats (AMR,

FLAC, M4A ,MP3 ,MP4 ,Ogg ,WebM and WAV) can be uploaded to Reveal for AI-powered transcription ■ In addition a Zoom recording url can be entered and text can be pasted into the platform ■ Already transcribed material can also be uploaded in .docx or .txt format ■ Reveal recognises differently prepared transcripts in terms of how speakers and timestamps are formatted, from which it will identify speakers ■ By

default Reveal supports 11 common languages of data (English, Dutch, Finnish, French, German, Italian, Polish, Portuguese, Russian, Spanish and Turkish). There is the ability to support an additional 70 languages which users can request to be added to their account at any point ■ When uploading data, sensitive information in the form of Protected Health Information (PHI) or Personal Identifiable Information (PII) contained within files can be redacted if required.

Study set-up to focus AI-generated responses

When setting up a study the user can optionally add contextual information to guide subsequent AI-generated responses ■ These include: one or more broad or specific research questions, hypotheses about which Reveal will look for supporting and contradictory evidence, a "persona" (e.g. "a qualitative researcher", or "a UX researcher", etc.), an analysis framework (currently 'thematic analysis', 'phenomenology', 'narrative analysis', 'case study', 'content analysis', 'discourse analysis', 'ethnography', 'IPA', Framework analysis', 'critical discourse analysis', 'hermeneutics', 'feminist theory', 'Actor-Network Theory', and Constructivist Grounded Theory') ■ The platform will also suggest potentially relevant questions based on the content of the uploaded materials ■ There are currently four choices for output language (English, French, German, Spanish) which does not need to the be same as the language the data are in. Study information can be updated at any point.

Figure 1. Observation Grid display in Reveal
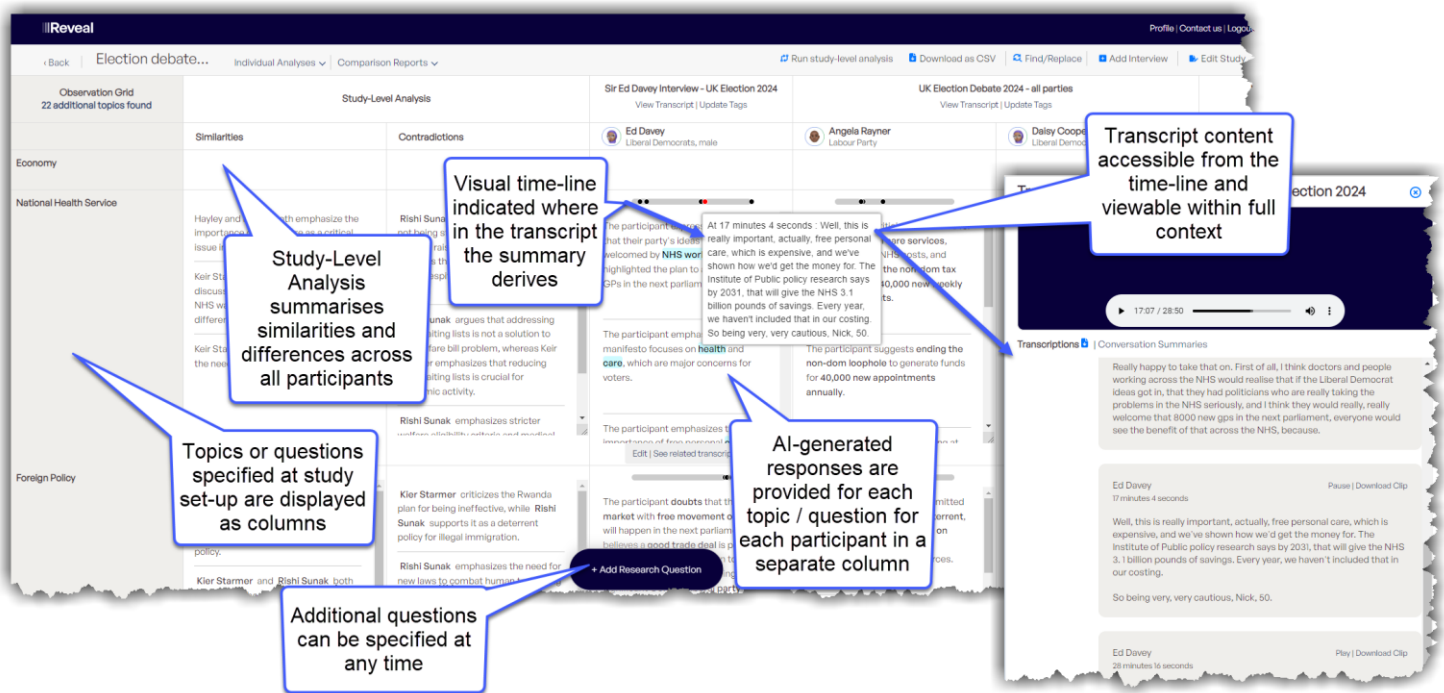
Automated transcription in Reveal

Transcription begins automatically and users are emailed when it has completed ■ Other work can continue within the platform while transcription is being performed ■ Users specify how speakers are named for display purposes and indicate which speaker(s) are interviewees ■ The transcribed audio/video can be played back from a particular speaker section within the platform, accessed from the textual transcript, and audio/video clips can be exported ■ As part of the transcription process Conversation Summaries are generated. These are mini-summaries of topic-based conversations within transcripts, bringing the context of questions and responses together. Each 'conversation' is attributed to the topics discussed.

Analysis tools in Reveal

Once a Study is set up, Reveal automatically generates a series of Individual Analyses and Comparison Reports ■ There are five types of **Individual Analyses**, focused on AI-generated responses to user-specified questions retrieved and displayed in relation to individual participants ■ In addition, users can build their own **Comparison Reports** to compare data in relation to study-level questions according to participant tags (e.g. socio-demographic characteristics) ■ Similarities and Differences between tagged cohorts are generated in Comparison Reports as well as comparing participants based on tags.

Figure 2. Observation Map display in Reveal

Individual Analyses in Reveal

The default display is the Observations Grid – a table containing the research questions specified at project set-up (displayed in rows) by participants (displayed as columns). Each cell in the grid is automatically populated with an AI-generated summary based on the intersection, which is connected to the source transcript. Where several extracts are drawn upon in generating the summary, this is visually indicated within each cell on an interactive timeline display, and each original extract can be displayed as well as the transcript reopened to view in context. Summary findings in the grid view can be edited. In addition, displayed adjacent in the Observation Grid is a Study-Level analysis of similarities and differences amongst participants for each research question ■ The Observations Map presents summarized responses to research questions in a visual display of topics and sub-topics, aimed at helping the user identify key themes ■ Hypothesis Testing is based on user specified hypotheses created for the Study (either when setting up or at any point subsequently). Reveal searches each transcript for evidence relation to each hypothesis and displays the result in a grid, indicating using whether there is or isn't evidence, or if they hypothesis is inconclusive ■ Exploratory Questions allow additional questions to be asked of the data within a study separately from any research questions specified in the study set-up ■ Quotes Finder will extract quotes from transcripts based on topics specified

Closeness to data and interactivity

At all stages of work within Reveal the user can access the underlying transcripts upon which the AI-generated content is based. For example, once audio/video has been transcribed clips can be played back to check accuracy and correct if necessary ■ In addition, every analysis display enables the underlying transcript to be accessed, to sense-check in context ■ Several displays additionally have a visualisation of the transcript indicating where the evidence is located (e.g. near the beginning of a transcript, or elsewhere) ■ These visualisations are also interactively connected with the underlying data, that can be pulled up in a pop-up within the analysis space, or accessed within source transcript to view the full context.

Figure 3. Quotes Finder in Reveal
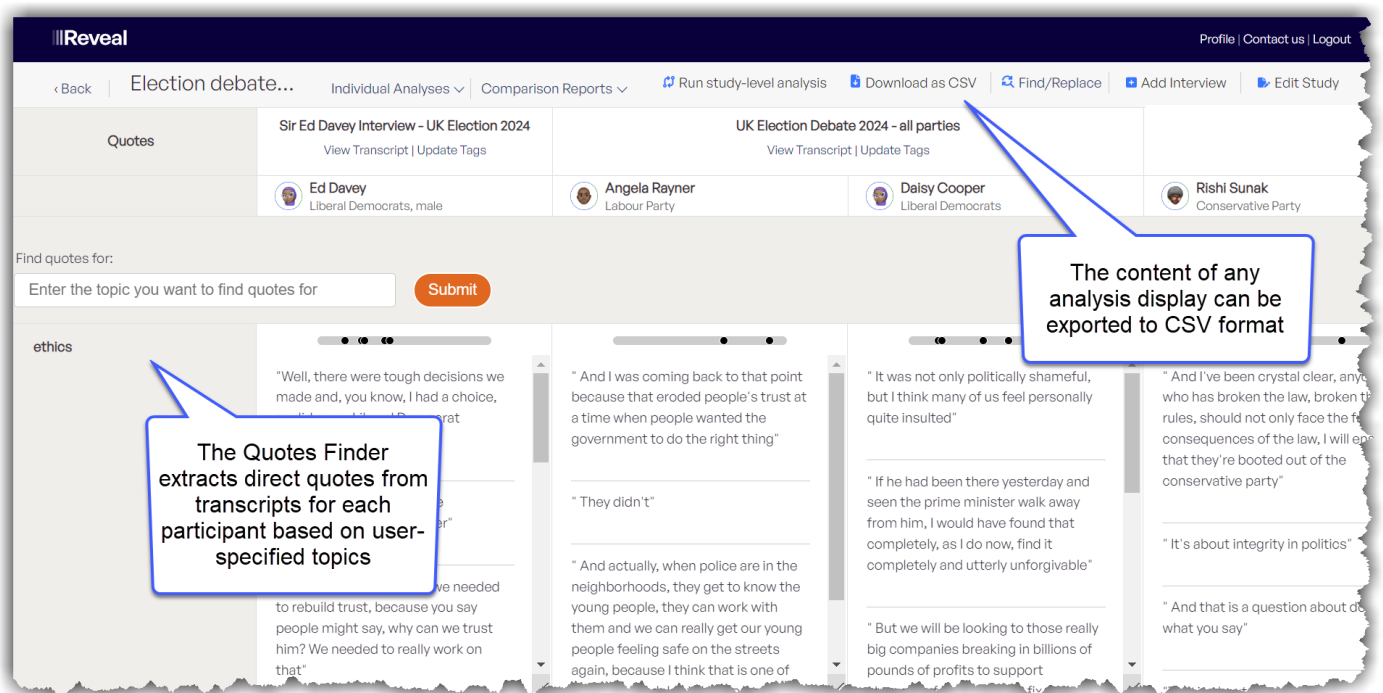
Data Organisation in Reveal

Factual characteristics about transcripts and participants is enabled through applying tags ■ As many tags as required can be associated with each participant ■ Comparison Reports use tags to compare participants

Writing Tools in Reveal

There are currently no spaces for the user to make notes about the analysis separate from transcripts ■ However there is a find/replace tool for making alterations to transcripts if errors are identified at any point ■ In addition, AI-generated summaries for participants are fully editable, allowing to users to add notes and edit / remove AI-generated content.

Output from Reveal

Any of the responses generated can be outputted into csv format, displaying for example the comparison

Comment on Reveal

- *__Reveal offers a range tools underpinned by Generative-AI__ designed for the analysis of text transcripts of interviews and focus-groups that are useful for exploring qualitative data.*

- *__Oriented towards the needs of market research__ there are some limitations for researchers working in other contexts, but those with any type of conversational qualitative data interested in using Generative-AI to support analysis will be interested in exploring what Reveal has to offer.*

- *__The interface is modern, clean and straightforward__ to navigate and it is easy to get started using the platform.*

- *__The ability to redact sensitive information__ at the point of transcription is useful, although more information on the models used to identify Protected Health Information (PHI) and Personal Identifiable Information (PII) make this feature more useful, and the ability to specify sensitive terms to redact would be valuable.*

- *__There are a variety of analysis tools and comparative representations__ which provide useful ways to interrogate for different purposes – for example making comparisons on the basis of participant characteristics, exploring the prevalence of topics / subtopics, and extracting quotes.*

- *__The ability to focus the AI__ around the use of personas and analytical frameworks as well as study objectives and hypotheses provides a degree of flexibility. Researchers will need to experiment with these possibilities to evaluate their utility in different methodological contexts.*

- *__There is excellent connectivity between AI-generated responses and the underlying source transcripts__ from individual summarized responses in all the displays. In particular the visual representation of where within a transcript a summarized response derives provides a useful high-level overview of the sequencing of topic-based conversations within each transcript. This is helpful and reassuring for the researcher who needs to sense-check and consider context. The ability to regenerate responses is useful when using these tools to familiarize and explore content.*

- *__The ability to edit responses within the platform__ is useful, to incorporate human interpretations with AI-generated summaries and responses to topic and question-prompts. However, the user must remember to format their additions to be able to track and differentiate AI-generated content from human interpretations.*

- *__The website contains useful information__ about Reveal's features, but the lack of a central help menu or searchable set of user instructions may be frustrating for some users. Although information on data security is provided, this lacks detail, although researchers can enquire for more explanation – __privacy@doreveal.com__ It would be helpful to users considering the use of generative-AI for qualitative analysis for there to be more transparency about how the tool works.*

- *__The tools are being continually developed and refined__ providing users with additional ways of harnessing generative-AI for qualitative analysis. Users will likely find it useful for additional data formats such as spreadsheets that contain demographic data and closed question answers to be importable.*

- *__Team-working__ is not currently enabled other than by sharing outputs, enabling multiple researchers to log into the same project would extend possibilities for those working collaboratively.*