AILYZE: 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 AILYZE 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 the software developer website. This document reviews AILYZE as of May 2024. Many thanks to James Goh for reviewing this document for accuracy.

Background https://www.ailyze.com

AILYZE is an online AI-driven qualitative analysis platform that uses large language models (LLMs) to autonomously interview stakeholders in any language, and to facilitate the analysis of docx and pdf text data in any language by generating thematic analyses, frequency analyses, summaries and responses to user-specified questions ■ Founded in 2023 by James Goh, in Boston, United States. Founder has a background in development economics and data science at MIT and The Wharton School ■ Three versions of AILYZE ■ ***AILYZE Lite*** is a free web-based service allowing users to upload up to three text documents (in either docx, pdf, xlsx, csv, xls format and with a limit of 10MB per file) and conduct four types of analysis (see below) ■ ***AILYZE Pro*** includes all the features of the Lite version for up to 50 documents plus additional features for more nuanced analysis (see below) ■ ***AILYZE Enterprise*** is a service provided by AILYZE that enables larger datasets to be custom analyzed using more powerful models than are available in the Lite and Pro versions (see below) ■ AILYZE functions using proprietary models that are trained on publicly available evaluation-focused data, including evaluation reports, interview transcripts, textbooks and research papers

Minimum System Specifications (recommended by developer)

AILYZE works with any desktop browser (e.g., Chrome, Safari, Firefox) for Basic / Advanced Analysis features and any desktop or mobile browser for the AI Interviewer feature.

Data security in AILYZE https://www.ailyze.com/data-security

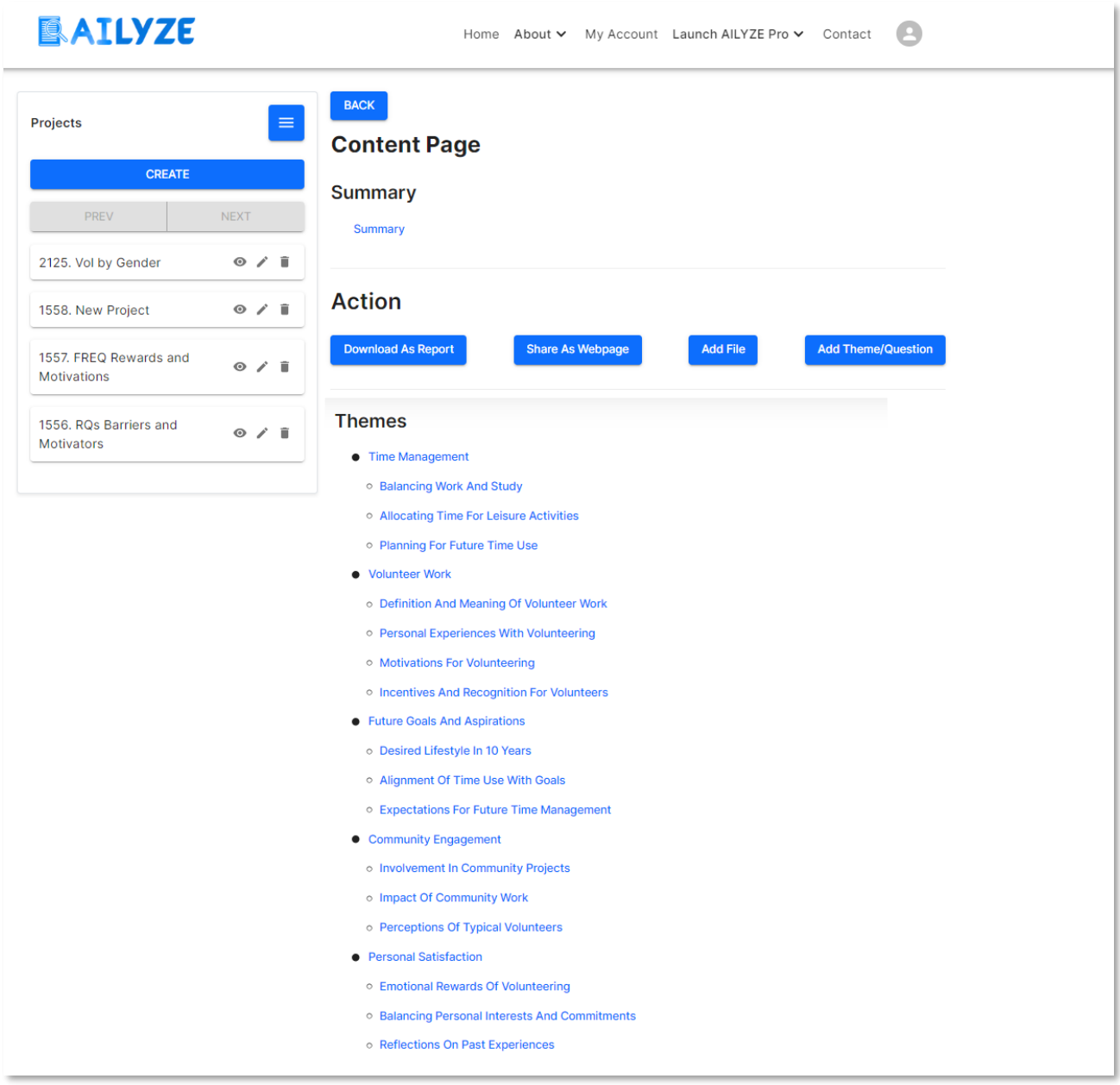
AILYZE complies with GDPR and CCPA regulations. Users can access HIPPA compliance by entering into a Business Associate Agreement with AILYZE ■ When using the Lite version uploaded data is deleted once the internet browser is closed ■ When using the Pro or Enterprise versions uploaded data is held until users specify to have it deleted ■ In both instances, data is encrypted at rest and in transit using FIPS 140-2 compliant 256-bit AES encryption, such that AILYZE cannot view uploaded data ■ Uploaded data is stored in a data center that is ISO-27001, SOC 1 and SOC 2-compliant ■ The service operates entirely over HTTPS, with API and application endpoints TLS only (v1.2) ■ Uploaded data are not used for training models.

Features in AILYZE Lite

Having uploaded up to 3 text documents of no more than 15,000 words per file, it is possible to choose between four types of analysis in AILYZE Lite (for no cost) ■ Summarize will generate a separate summary of each text document based on the criteria specified (Essay or Bullet Points / Language of Response (all 100+ major languages) / Optional additional Instructions (specified by the user in a prompt) (e.g., elaboration of acronyms, context of the evaluation, format of the response desired). Summaries are provided within the browser and can be copy/pasted out. Choose between cross-document summary or individual summaries for each document ■ Ask a specific Question requires the user to specify a question that the tool uses to generate a response. Multiple questions can be asked to delve deeper into the documents. The responses are provided within the browser, with 'footnotes' that include direct quotes from the documents illustrating from where the response derived, and links to the original source to view the quotes in context ■ Conduct Thematic Analysis will generate thematic summaries of uploaded documents

with each theme contained within a separate paragraph. Thematic summaries can be focused by adding optional instructions and the length of responses specified, and can be copy/pasted out ■ Compare Viewpoints across Documents allows the user to ask questions of the data and for the responses to be provided comparatively based on user-specified characteristics at the document level (e.g. demographic information about participants per uploaded transcript)

Figure 1. The main Interface showing list of Thematic Results in AILYZE Pro

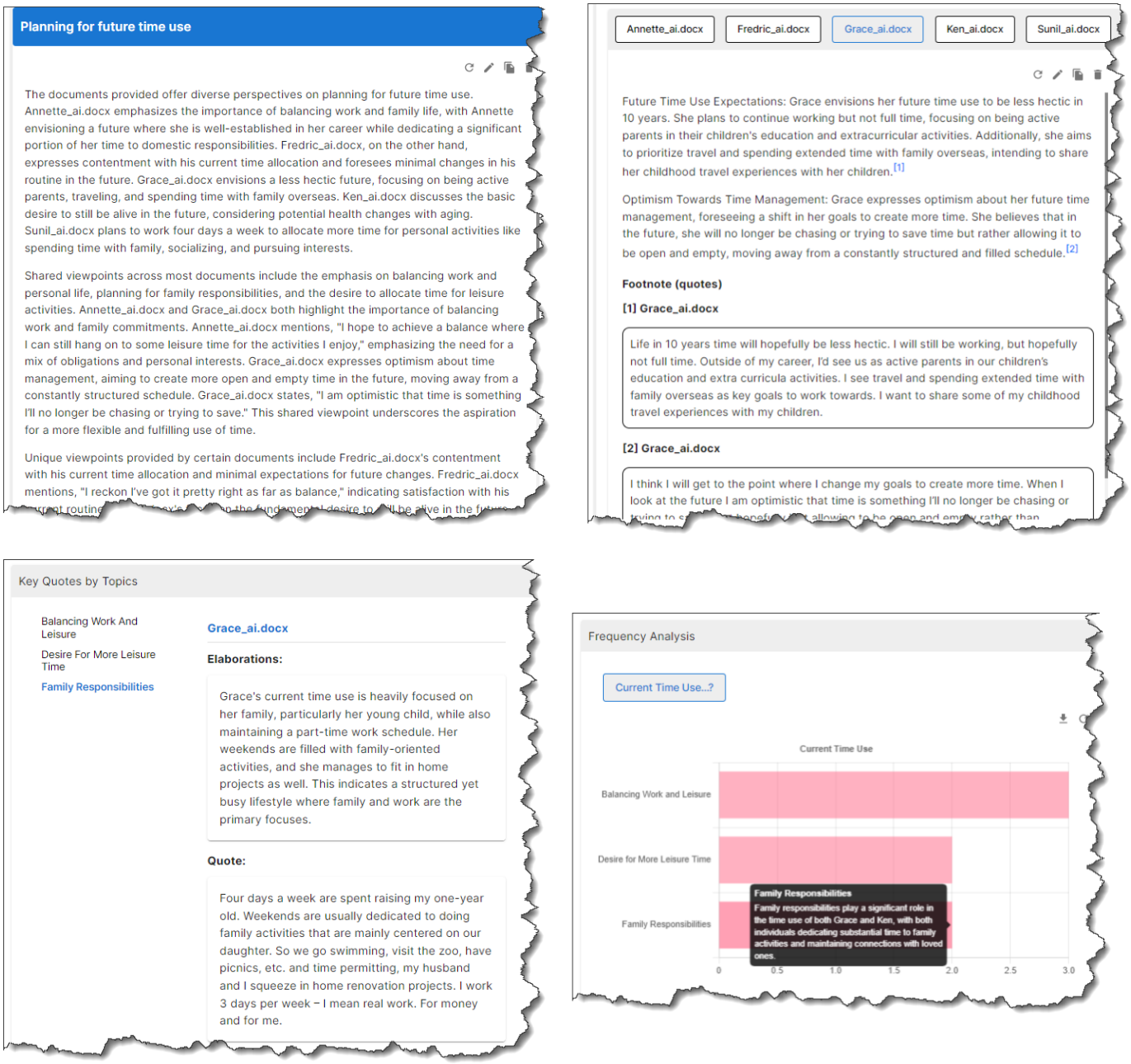


Features in AILYZE Pro

In addition to the basic analysis features available in the Lite version (as described above), AILYZE Pro provides advanced analysis features for a monthly subscription fee (see website for up-to-date costs) ■ Analysis of up to 50 documents containing up to 90,000 words per file ■ Advanced Thematic Analysis options include specifying separate research questions to focus the analysis, the ability to choose which documents to compare viewpoints across based on categories (e.g.

demographics or other document-based features), and (currently experimental) frequency analysis of viewpoints ■ After specifying the focus the project is sent for creation which can take up to an hour – the result is provided in a navigable structured list format, allowing responses to the different elements to be accessed and explored within the browser. See https://www.ailyze.com/dummy-project for an example of the AI-generated report ■ Results are provided based on the research questions asked, thematically (across documents) and per document. Additional questions can be asked from the results to further explore the data, and additional documents can be added to include in the analysis. The results can be shared as a webpage enabling other to view and navigate the results but not to edit the project, or downloaded as a Report (.docx format).

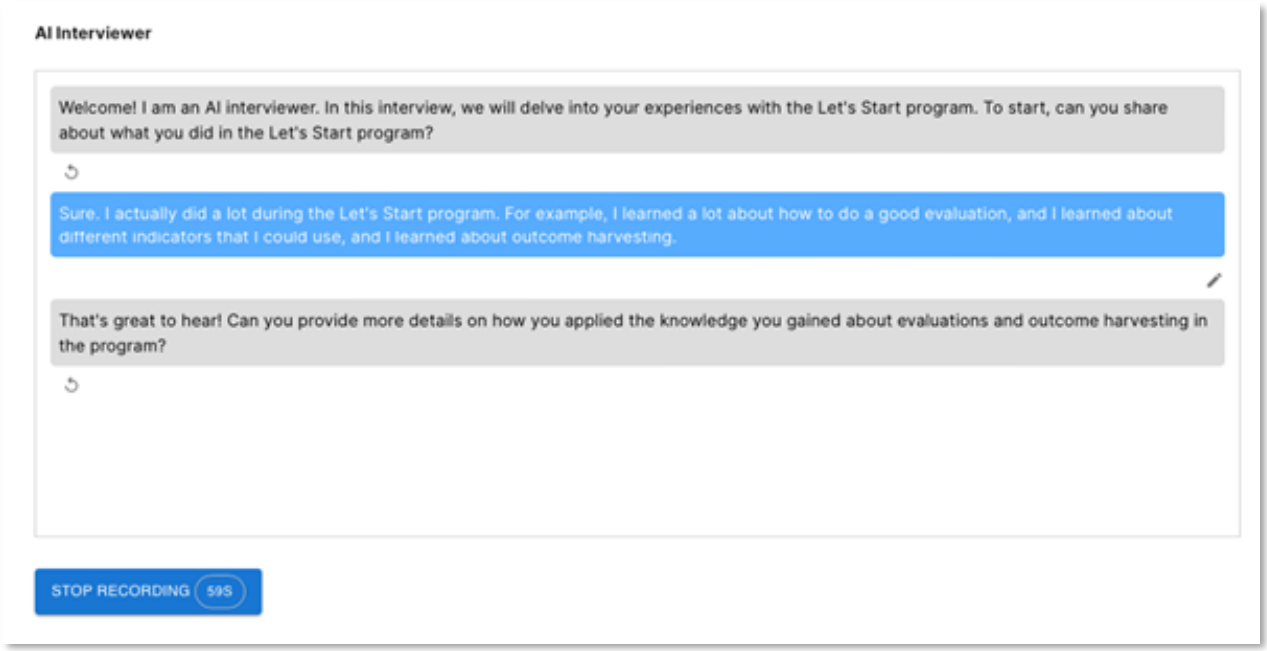
Figure 2. Selected results of Thematic Analysis in AILYZE Pro



Features in AI Interviewer

The model is trained on publicly available interview transcripts that were conducted by qualitative researchers using multiple languages[1]. The AI interviewer generates contextually relevant replies and further prompts the interviewee based on their responses, and asks clarifying questions when interviewees provide ambiguous or unclear responses, for example asking for examples, seeking clarifications and probing for underlying assumptions

Figure 3. AI interviewer



The Structure of work in AILYZE

The AILYZE interface and the structure of work differs slightly depending on whether you're working with the Lite or Pro version, but in both analysis proceeds from the initial selection of text files to interrogate ■ In the Lite version everything happens within a browser, with features selected sequentially down the page: starting with the uploading of files, then choosing the type of analysis and specifying the parameters and optional instructions, etc. The results are subsequently displayed on the same page ■ The Pro version has an additional side panel where projects can be created and managed

Data types and format

Data files are referred to as Documents in AILYZE ■ Text files in the following formats are uploaded for AI-assisted analysis : docx, pdf, xls, and xlsx ■ When uploading spreadsheet data the type of analysis undertaken can either be to conduct thematic analysis based on text in a column or to categorise text in each cell in a column ■ Text in all 100+ major languages can be processed, and responses can also be provided in a range of languages ■ The AI Interviewer tool can also speak and understand a variety of languages

[1] Current languages are: Afrikaans, Arabic, Armenian, Azerbaijani, Belarusian, Bosnian, Bulgarian, Catalan, Chinese, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, Galician, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Indonesian, Italian, Japanese, Kannada, Kazakh, Korean, Latvian, Lithuanian, Macedonian, Malay, Marathi, Maori, Nepali, Norwegian, Persian, Polish, Portuguese, Romanian, Russian, Serbian, Slovak, Slovenian, Spanish, Swahili, Swedish, Tagalog, Tamil, Thai, Turkish, Ukrainian, Urdu, Vietnamese, and Welsh

Closeness to data and interactivity

In AILYZE the analysis is done by the model and therefore users are not working with the source text in its original format directly themselves ■ Nevertheless, most responses generated are linked to the source text from which they derived via citations and footnotes, which can be clicked on to view where the quotes are located within the broader context of the document. Therefore, users can access the original source material to retrieve context, and to assess the interpretive value of the generated responses

Data Organisation in AILYZE

There are currently no special features for organizing factual features in documents in AILYZE via for example numeric or categorical variables ■ However, comparisons on such bases can be undertaken in the Pro version through specifying questions accordingly, through the use of instructions for these purposes, and using the option to compare viewpoints across groups (e.g., user will specify which files are male/ female and the analysis will compare viewpoints across these groups accordingly) and to conduct a frequency analysis of viewpoints (e.g., how many documents expressed a certain viewpoint).

Writing Tools in AILYZE

There are currently no specific note-taking or other writing spaces within AILYZE although responses generated can be edited

Searching and interrogating the dataset in AILYZE

All searching and interrogating in AILYZE happens by asking the model questions

Output in AILYZE

The results of analysis undertaken in the Pro version can be shared as a navigable webpage or downloaded as a text file (docx). In the Lite version, results can be extracted from the platform using simple copy/paste.

Team working in AILYZE

AILYZE is a single-user platform but analysis projects created in the Pro version can be shared with others in the form of exports (navigable webpage or downloaded report).

Comment on AILYZE

- ***AILYZE offers a range of generative-AI tools*** *designed for the analysis of text documents that are useful for exploring qualitative data. It is easy to get started with the Lite version which provides those interested in this genre of analytic tools quick and easy access to experiment. Generally the platform is easy and straightforward to operate. The Pro and Enterprise versions extend analytic possibilities providing additional ways of questioning uploaded qualitative materials in more depth and comparatively.*
- ***There is good connectivity between AI-generated responses and the underlying source texts*** *from most results, although this is not always present – for example from the summaries. The ability to regenerate responses is useful when using these tools to familiarize and explore content.*
- ***The ability to edit responses within the platform*** *as well in exported results is useful, to incorporate human interpretations with AI-generated summaries and responses to question-prompts. However, the user must remember to format their additions to be able to track and differentiate AI-generated content from human interpretations. It would also be helpful to many users to be able to track questions asked and analyses undertaken more easily within the platform.*
- ***The website contains a lot of useful information*** *about the features and potential applications of AILYZE in a series of FAQs, 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. 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.*