5-2024

# Streamlined workflow for qualitative data analysis with Whisper and ATLAS.ti

Bella RATMELIA

Bryan LEOW

Danping DONG

# Streamlined Workflow for Qualitative Data Analysis with Whisper and ATLAS.ti

**Bella Ratmelia** – Senior Librarian, Research & Data Services

**Bryan Leow** – Senior Librarian, Research & Data Services

**Dong Danping** – Senior Librarian, Research & Data Services

SMU Libraries

16 May 2024

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Outline

1. Qualitative Research Workflow and considerations
2. Transcribing with Whisper API from OpenAI
3. Qualitative Coding with Atlas.ti
4. Readings and Resources

# Types of Qualitative Research

# A typical workflow for interview-based research

## Conceptualization

Literature review

Research design

## Data collection

Conduct interviews and record

## Data processing

Manual transcription

## Data analysis

Data coding & thematic analysis

# A more efficient workflow: with QDA software & AI tools

## Conceptualization

Literature review

Research design

## Data collection

Conduct interviews and record

## Data processing

AI transcription

## Data analysis

Computer-assisted QDA software, AI coding, visualization,

& thematic analysis

# Case Study – How Students Use AI tools for Learning

- Libraries wants to find how students respond to AI tools for the purposes of learning
- Objective is to see how we can integrate the tool into our workshops and research consultations, while keeping in context SMU's overall strategy on the use of AI tools
- Libraries Team intends to look at the following

    i) Focus groups of SMU users who may or may not have used AI tools Academic Literature which have

    ii) Academic literature on the topic

- What tools can help us?
    - **Whisper** for transcribing the focus group sessions
    - **ATLAS.ti** for collating and analysing all the data we have collected

# Transcription with Whisper API

# What is Whisper API?

OpenAI's state-of-the-art open-source speech-to-text model that can be used to:

- **Transcribe** audio into whatever language the audio is in.
- **Translate** and **transcribe** the audio into English.

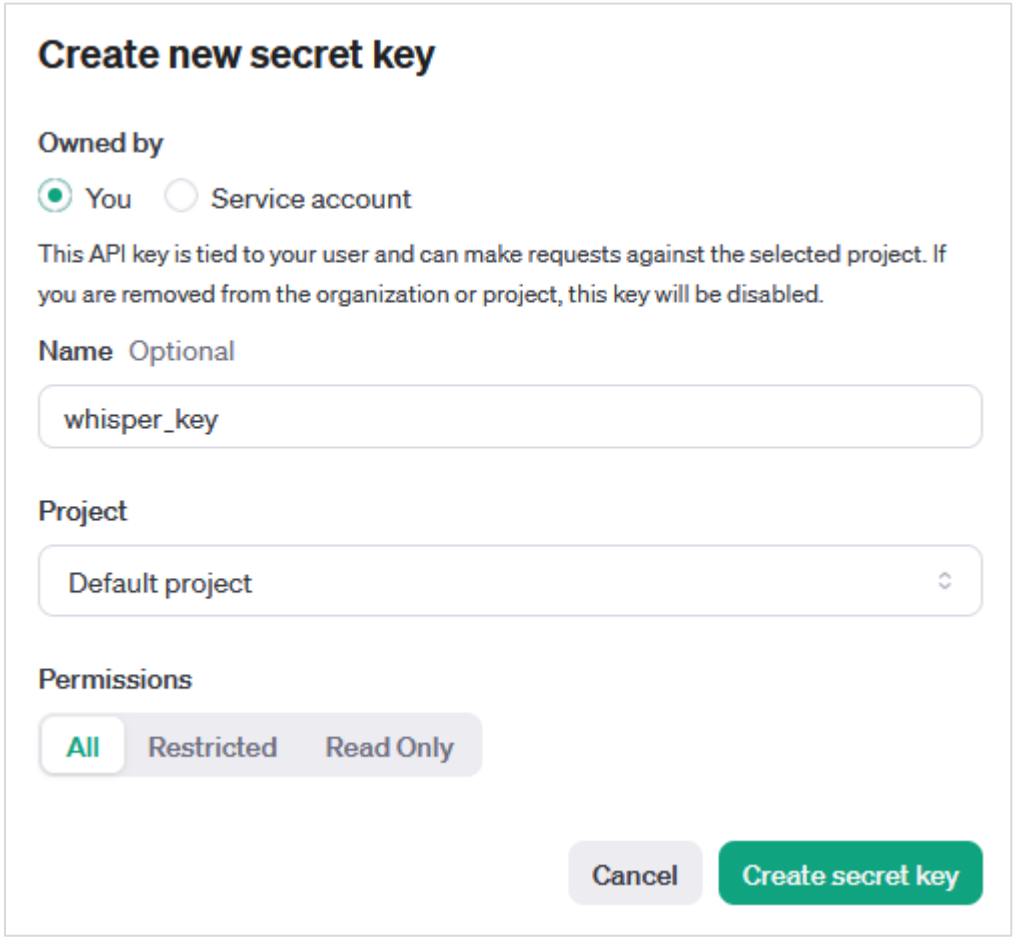
Available through **API access** (not to be mistaken with ChatGPT premium!) – **some coding is required to use Whisper**.

## API Requirements:

- Cost: $0.006 / minute (rounded to the nearest second)
- File types supported: mp3, mp4, mpeg, mpga, m4a, wav, and webm.
- File uploads are limited to 25 MB per API call (roughly ~20 minutes worth of recording in an .m4a file)
- Audio files must be at least 0.1 second long

# Getting the API Key

1. Login to https://platform.openai.com/api-keys
2. Click on “Create new secret key”, and you will see a popup that looks similar to the screenshot on the right.
3. Once you click on “Create secret key”, an API key will be displayed.
4. **Immediately copy and save this ID into the environment variables or somewhere else safe! You won’t be able to view this key again.**
5. Use this key in your API calls.



# Transcription workflow: Single Speaker


**Step 1:** Convert audio to mp3 in 128kbps bitrate to compress size.

- Not a must, but I would recommend it!

**Step 2:** if your audio is large, slice the audio files to smaller chunks (10-15mins long each) with pydub

**Step 3:** Transcribe the audio files one by one with **Whisper API**

- Test the transcription one 1 chunk first to check the quality.

**Step 4:** Combine the transcription files

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Sample .txt Transcription Output

audio-sample-focus-group-1.txt

Okay. Okay. Hi, everyone. My name is Bella. Nice to meet you and thank you for being here today for this focus group session. So, just to reiterate, the objective of the session today is to gather insights into how novice learners are making use of chat GPT or other AI tools to support their coding endeavors. As you may know, SMU libraries offer Python and R classes, so we are particularly interested in exploring how AI tools like chat GPT, Copilot, and others can be meaningfully integrated into our coding workshops. So, this session will take about an hour or so, probably less. As much as possible, I will ensure that your comments are kept confidential. It will be anonymized such that even if we quote you in the final report, it will not be traced back to you. Please feel free to let me know if there's anything that you don't want to be recorded at all. But having said that, please don't hold back judgment and feel free to be as descriptive as you can. As a reminder, this interview is entirely voluntary on your part. If for any reason at any point in time during this focus group session you want to stop, just let me know and we will end the session. There will be no penalty whatsoever and let me know if you'd like to discard everything that you have told me until that point. Do you have any questions for me? No. Okay, so let's begin. We'll start with some warm-up questions. So, could you tell me about the coding work that you do or you're trying to learn? Okay, for me, recently I'm trying to use Bird Topic which is a tool to cluster topics. So, that's one of the things I'm trying to do. I have a dataset that contains a lot of chat transcripts and I want to uncover the topics that's present in

# Pros and Cons

## Pros

- Recognize Singlish pronunciation fairly accurately.
- Multilingual support
- Relatively cheap, as it charges on a per minute basis.
- Open-source (i.e. free version) available on GitHub

## Cons

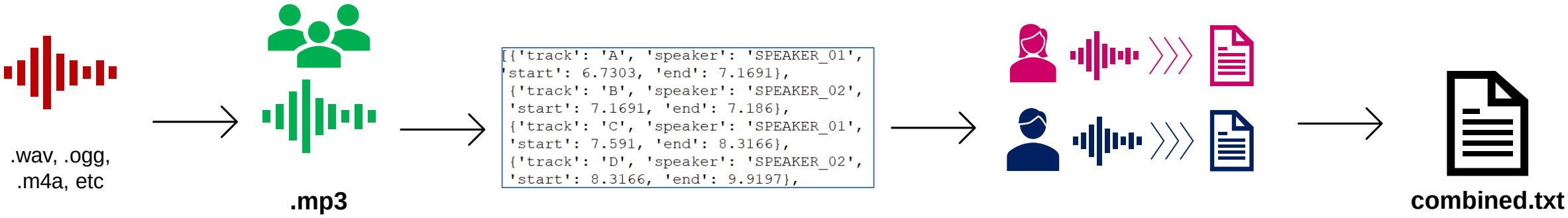
- Requires a little bit of coding.
- Depends on internet connection.
- May not support some languages robustly.
- **Does not distinguish between speakers, if there are multiple speakers**.
  - This could be problematic as most interview and focus group recordings will undoubtedly have multiple speakers!

# Speaker Diarization for Audio with Multiple Speakers

Speaker diarization: The process of partitioning an audio stream containing human speech into homogeneous segments according to the identity of each speaker. (Wikipedia)

- Currently not built-in to Whisper (yet)
- Popular model for diarization: **pyannote.audio**, available through Hugging Face

# Transcription workflow: Multiple Speakers




**Step 1:** Convert audio to mp3 in 128kbps bitrate to compress size.

- Make sure to convert to mp3 or wav, as pyannote does not support .m4a
- If your audio is very long, you may want to slice them into smaller parts of about 3-4 minutes each.

**Step 2:** Diarize audio with **pyannote** and save the speech turn with **start** and **end** timestamp info as a list or json.

- The diarization part can take a while even for small files (a 3-mins .mp3 took 15 mins of diarization), so be sure to save this info to be used later!

**Step 3:** use this **start** and **end** timestamp to slice the audio file with pydub and transcribe the speech with **Whisper API**.

- You can save each transcription directly to a txt file, or just update the list containing the speech turn timestamps with the transcription.

**Step 4:** combine and tidy up the transcription into a single file

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Sample .txt Diarization + Transcription Output

diarizer_combined_transcript.txt

```
17  SPEAKER_00 :
18  SPEAKER_00 : This is a conversation about using ChatGPT to learn Python spoken in English.
19  SPEAKER_00 : So, let's begin.
20  SPEAKER_00 : We'll start with some warm-up questions. So, could you tell me about the coding work that...
21  SPEAKER_00 : You do, or you're trying to learn.
22  SPEAKER_01 : Okay, for me, recently I'm trying to use BirdTopic, which is a tool to cluster topics.
23  SPEAKER_01 : So, that's one of the things I'm trying to do. I have a dataset that contains a lot of chat
    transcripts. And I want to uncover the topics that's present in this chat transcripts dataset. And another kind of
    work might be things like analyzing a dataset. I'm working on a survey dataset recently. So, that's some examples
    of the recent work I've been doing.
24  SPEAKER_02 : For me, I think it's more beginner-level stuff, so I'm also trying to understand how Python can be
    used to clean up data, or visualize data, or...
25  SPEAKER_00 : So, seems like mostly Python.
26  SPEAKER_02 : For me, Python.
27  SPEAKER_00 : For now.
28  SPEAKER_00 : For me, it's also mainly Python. I see. So, um... When did you start learning?
29  SPEAKER_02 : This is a conversation about using ChatGPT to learn Python spoken in English.
30  SPEAKER_01 : ..SPEAKER_02 : ..SPEAKER_01 : ..SPEAKER_02 : Oh, when was that? A month ago?
31  SPEAKER_01 : Mine is one year, two years. I mean, I kind of like touch on that, but I didn't really use a lot until
    recently, yeah.
32  SPEAKER_00 : Kind of like...
33  SPEAKER_00 : Okay, so...
34  SPEAKER_01 : This is a conversation about using ChatGPT to learn Python spoken in English.
```

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Common issues you may encounter (and how to fix them)

| Issues | Solution |
|---|---|
| .m4a files are not always readable by Whisper API, even though it is listed as one of the acceptable file format in the documentation. | Convert to mp3 or another format such as .webm or .wav. So far, .mp3 seems to be the sweet spot between size and audio quality. |
| The transcription seems to stop halfway. | Try slicing the audio file to smaller files; Whisper has a limit of about 25 MB per API call. |
| Whisper removes filler words (this can be an issue if you want a transcription that's as accurate as possible). | Add a prompt in the API call to instruct Whisper to include filler words. |
| Whisper inaccurately detect acronyms or jargons. | Adding the jargons or acronyms in the prompt will help Whisper to fix some of the inaccuracies. |
| Whisper "hallucinates" or churn out gibberish.<br>`So, I think, like, I don't know, like, I don't know, like, I don't know, like, I don't know, like, I don't know, like, I don't know, like, so, I think, like, like, it made me feel like, like, it made me don't feel like` | You can re-try the transcription for that specific segment, sometimes this will fix it (based on my experience). Adding a prompt to ask Whisper to be as close as possible to the spoken word sometimes helps! |

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# A unique issue for speech with accent

## Issue: Auto translation to non-English

This example is a transcription of a talk delivered in English by an Indonesian speaker.

Halfway through the transcription, Whisper “decided” to auto-translate the transcription to Bahasa Indonesia.

## Solution: Include a prompt in the API call that asks Whisper to transcribe the speech to English only. If this doesn’t work, you may need to adjust your prompt.

kamiliah-artist-talk.txt

301 There are words that exist in the Indonesian language that suggest a certain quality of time for example that I have yet to come across in English.

302 I think yeah I think you bring up like a really good point about you know not representing sort of like community or an experience to share your own personal like to share it from your own personal perspective because I think there's some in doing that I think there's an element of staying truthful to yourself which I think has the which actually has the effect of making your words a lot more accessible to other people because I think in bringing that personal element and people can really connect to your works on an also much more personal level as opposed to maybe like a more abstract objective like discursive level you know yeah which is I mean to me that's my experience with both your works yeah so I think we move like to be separate to time so that's something I actually do want to talk about as well I guess I think Bodhi with your works you yeah there is a lot of movement through time so movement to sort of like looking at experiences of your grandfather but also your father as well and even moving sometimes even further back with drawing on the myth you look at like this mythological origin story

303 Kemudian dengan kata-kata imortal, anda berpindah ke masa yang sangat dalam seperti 15,000 tahun yang lalu.

304 Apa yang anda bayangkan daripada perjalanan psikologik itu melalui masa? Saya merasa lebih terperangkap dalam cara yang aneh terhadap kewujudan.

305 Contohnya, setiap kali saya mengatakan perkataan berwarna hitam, yang merupakan warna kegemaran kami, saya mengatakan perkataan yang telah diatakan oleh seseorang 15,000 tahun yang lalu.

306 Ia membantu untuk mengarahkan saya ke dunia.

307 Pergerakan ini melalui masa, apa yang ia lakukan bagi saya ialah ia merupakan perubahan yang mempengaruhi kewujudan saya.

308 Perasaan pengetahuan tentang masa silam adalah sesuatu yang sangat sederhana.

309 Ia merupakan beberapa jenis tanggungjawab yang saya boleh menahan untuk memasukkan diri saya ke kehadiran.

310 Itulah cara saya mengalami pergerakan ini.

311 Terdapat banyak pergerakan dan pergerakan yang berulang-ulang.

# Resources

- Pyannote: https://github.com/pyannote/pyannote-audio or https://huggingface.co/pyannote/speaker-diarization-3.1
- Whisper: https://platform.openai.com/docs/guides/speech-to-text
  - API documentation: https://platform.openai.com/docs/api-reference/audio/createTranscription

# Other alternatives

1. **Text-to-Speech V2** by Google - https://cloud.google.com/speech-to-text/v2/docs
   - Also comes with speaker recognition feature
   - Comes with a separate model specifically for medical dictation/conversation
   - You can opt-in for data logging, which will make the transcription cost about 30% cheaper
   - More expensive than Whisper
2. **distil-whisper** via Hugging Face - https://github.com/huggingface/distil-whisper
   - Smaller than OpenAI's Whisper, but faster ("6x faster, 50% smaller, within 1%-word error rate" according to the description)
   - Only available for English speech recognition
3. **Pyannote.ai** - https://pyannote.ai/
   - Commercial version of open-source model pyannote.audio.
   - Additional costs on top of Whisper API (EUR 0.15 / hour)
   - Should provide more precise and faster diarization (in theory, at least!)

# Qualitative Coding with Atlas.ti

# Activate ATLAS.ti license (for SMU users only)

## Activate License

- Fill up the license request form
- Create an ATLAS.ti account using SMU email
- You will receive a link to activate the 5 concurrent users license provided by SMU Libraries

## Important Note

- **Save your projects & log out after each use**
- to prevent accidental data loss
- and to release the seat for other SMU users

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Investment & Data Studio (IDS)

https://researchguides.smu.edu.sg/ids

Located @ Level 3, Li Ka Shing Library

## Qualitative Data Analysis (QDA) software

- NVIVO (at the IDS only)
- ATLAS.ti – 5 concurrent users and can be accessed remotely

**Quantitative & Statistical Software**

- JASP
- Matlab
- Mathematica
- Microsoft Power BI Desktop
- OpenRefine
- SAS
- SPSS
- STATA
- Tableau Desktop - Lab License
- QGIS

**Qualitative Analysis Software**

- NVIVO 14 (only on Data Software Terminals D01 and D02)
- Atlas.ti (request for a license here - please use your SMU email to login and access the form)
- Covidence (email library@smu.edu.sg for access)
- ProQuest Text and Data Mining (TDM) Studio

**Programming**

- Anaconda
  - Python
  - NumPy/SciPy, MatplotLib, pandas, statsmodels, Scikit-learn and more
  - NTLK
  - Julia
  - Jupyter Notebook
- R/Rstudio
- Github Desktop
- MySQL Workbench
- Notepad ++
- Visual Studio Code

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Outline

- Introduction and overview of qualitative research
- ATLAS.ti Demo
  - Organize and code data
  - Analysis and visualization
  - Collaborate in teams

# Qualitative research

- Qualitative research at its core, ask **open-ended questions** whose answers are not easily put into numbers such as ‘how’ and ‘why’
- Qualitative research is looking for **themes** and **patterns** that can be difficult to quantify
  - i.e. the process of interpreting / analysing raw data
- Important to ensure that **context** is not lost by quantifying something that is not meant to be quantified

Source: National Library of Medicine, https://www.ncbi.nlm.nih.gov/books/NBK470395/

# Stages of Qualitative Analysis

**Preparation**: The first step is to choose the material or content to be analyzed. This could be interview transcripts, documents, media outputs, or any form of communication that is related to the research question. Researchers immerse themselves in the content to gain an overall understanding and develop initial impressions.

**Organizing**: During this stage, the material is broken down into manageable pieces, usually through coding. Each segment of the content that represents a single idea is assigned a code. Codes can be pre-defined, emergent, or a combination of both, depending on the research approach.

**Reporting and interpreting**: The final step involves analyzing the coded segments and looking for patterns or themes. These themes are then interpreted in relation to the research question and the broader context. This may include understanding implications, drawing conclusions, or formulating further theories.

# ATLAS.ti

A qualitative research tool for analysing large volumes of text, such as interview or focus group data

- Upload text documents in various formats
- Apply codes/categories to selected text sections
- Visualize your analysis in various forms, from reports, to charts or word clouds
- AI assistance for qualitative coding and code visualization

# Types of Qualitative Research

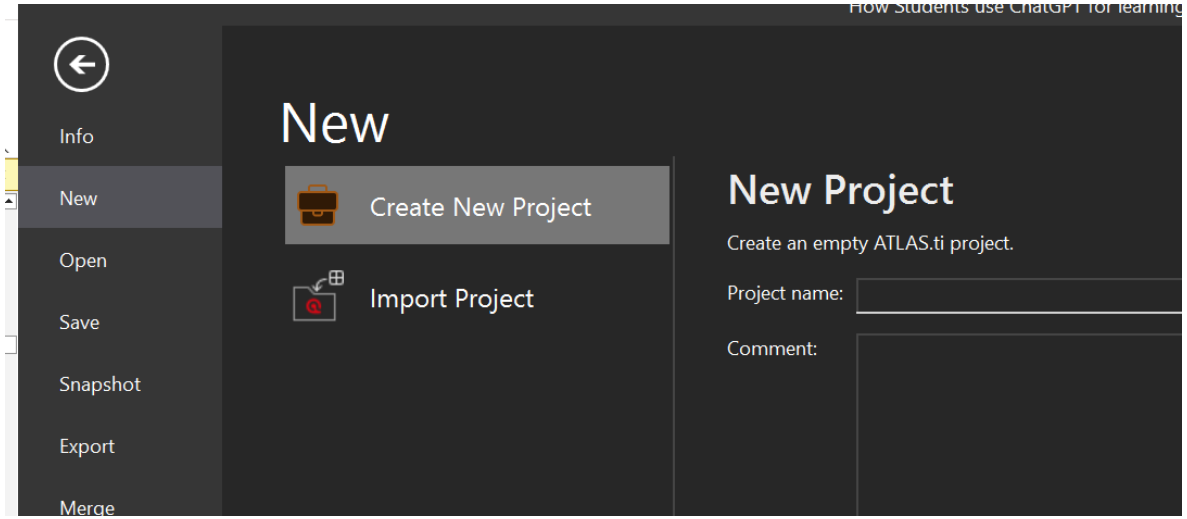
# Getting Started

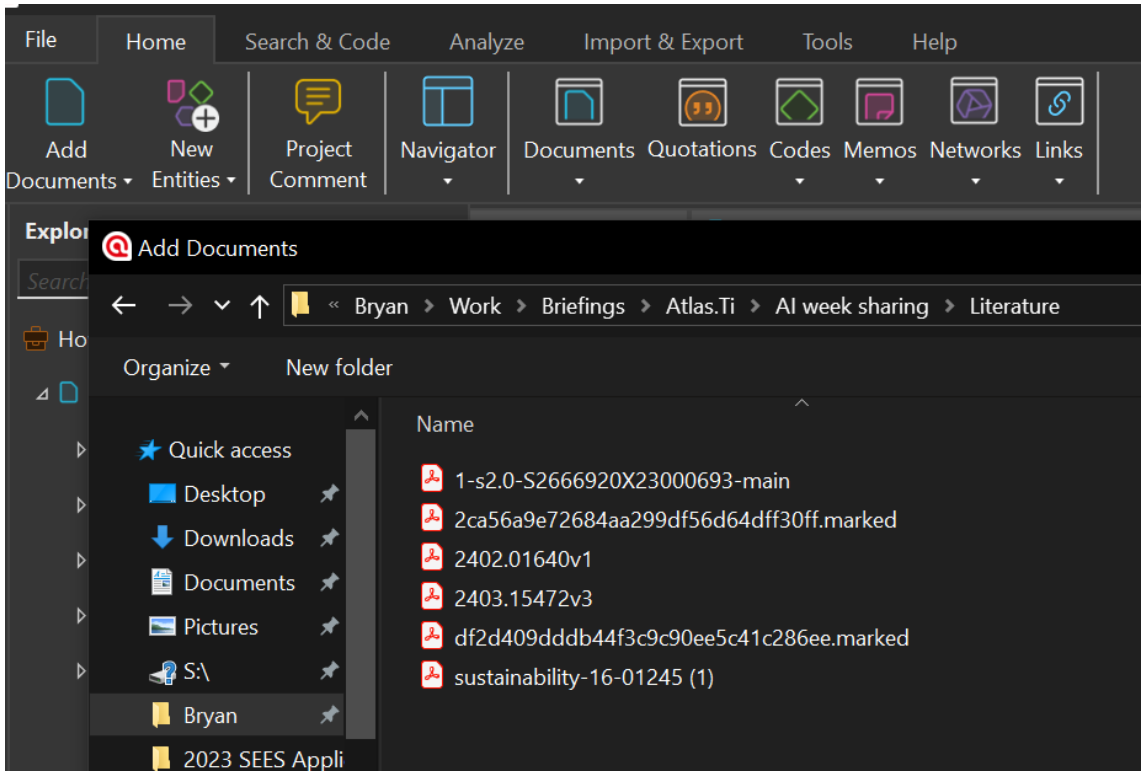
- Import files / survey data
- Understanding fundamentals of coding / quotes etc

**Download workshop files at**
https://tinyurl.com/486c3hb3

# Creating New Project and Importing Data

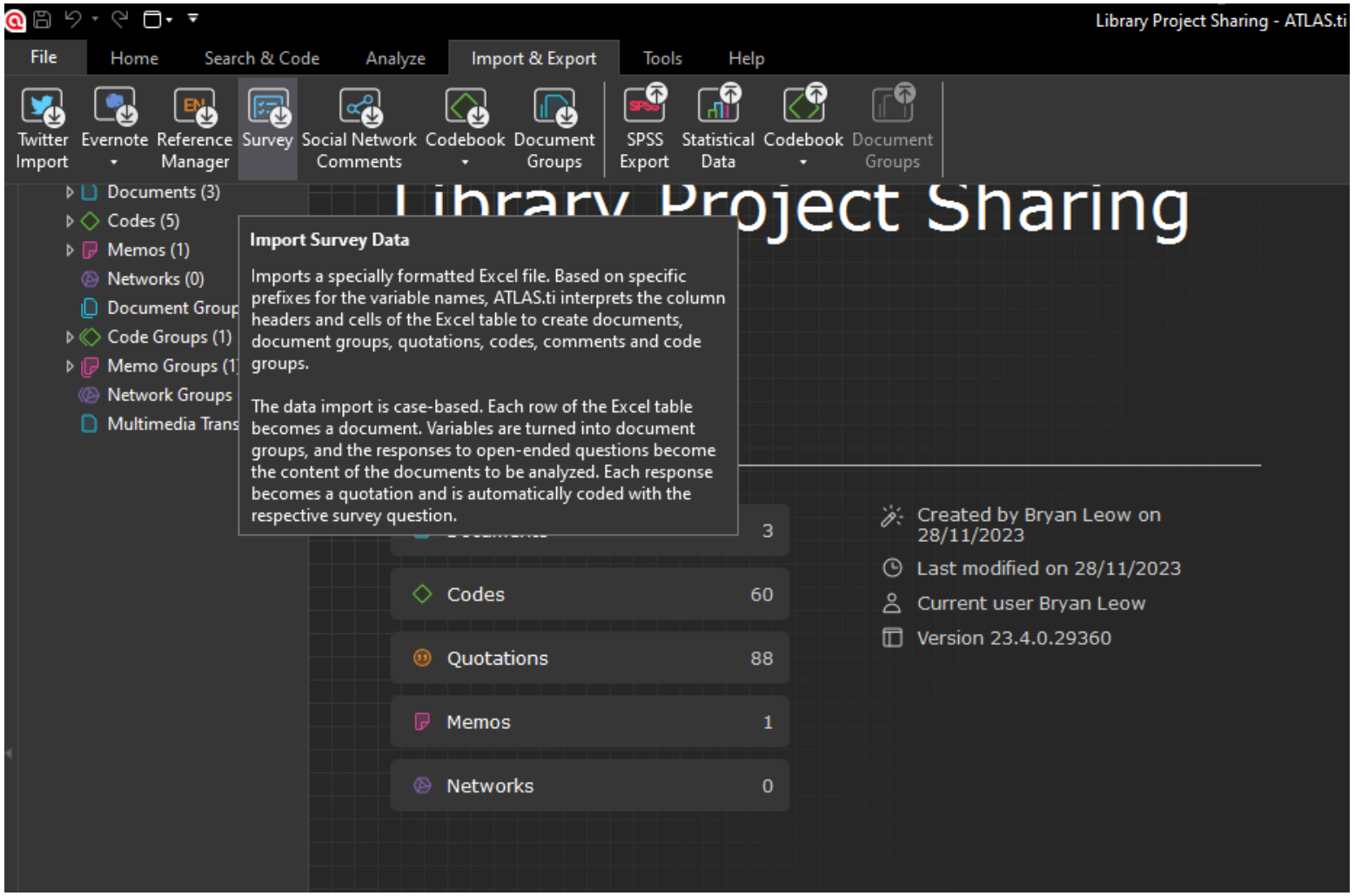
- Recall the study that Library wants to do – finding out how students respond to AI tools for the purposes of learning
- With Whisper, we have a set of transcripts of the focus group, and separately, the team has also done some searches and found literature that they want to review holistically.
- Whats next?



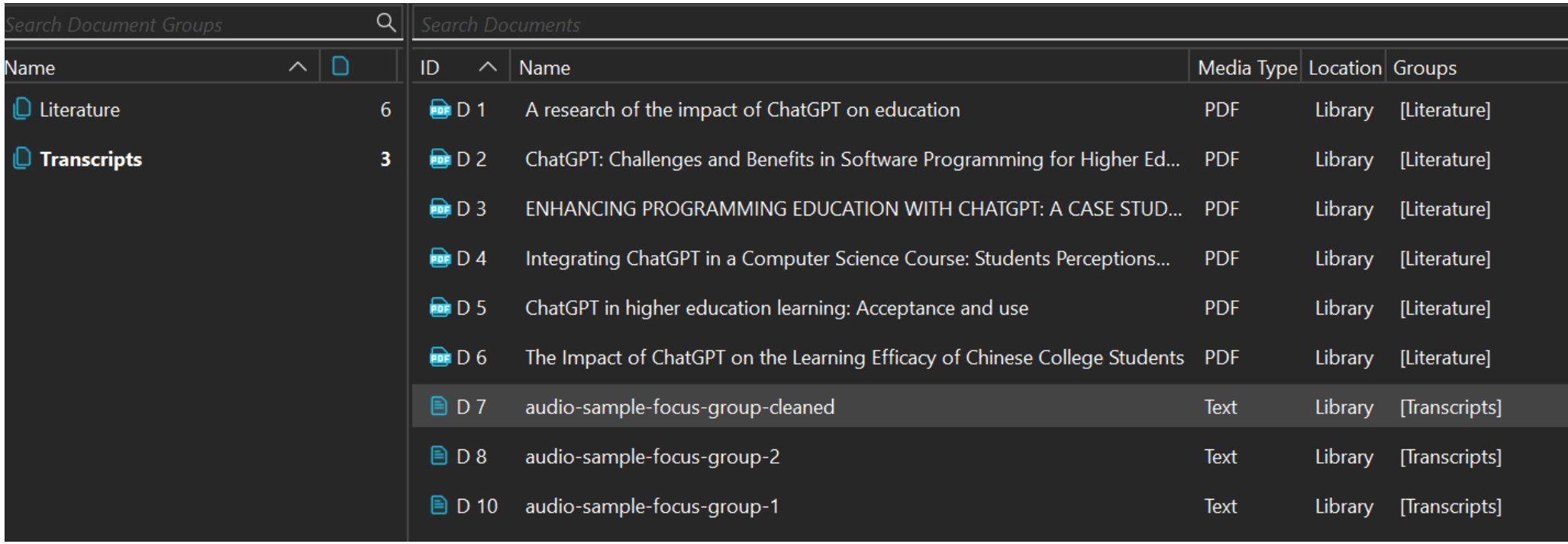
# Importing survey data

- Take note that your survey data will need to be clean and formatted to Atlas.Ti's (quite) exacting requirements



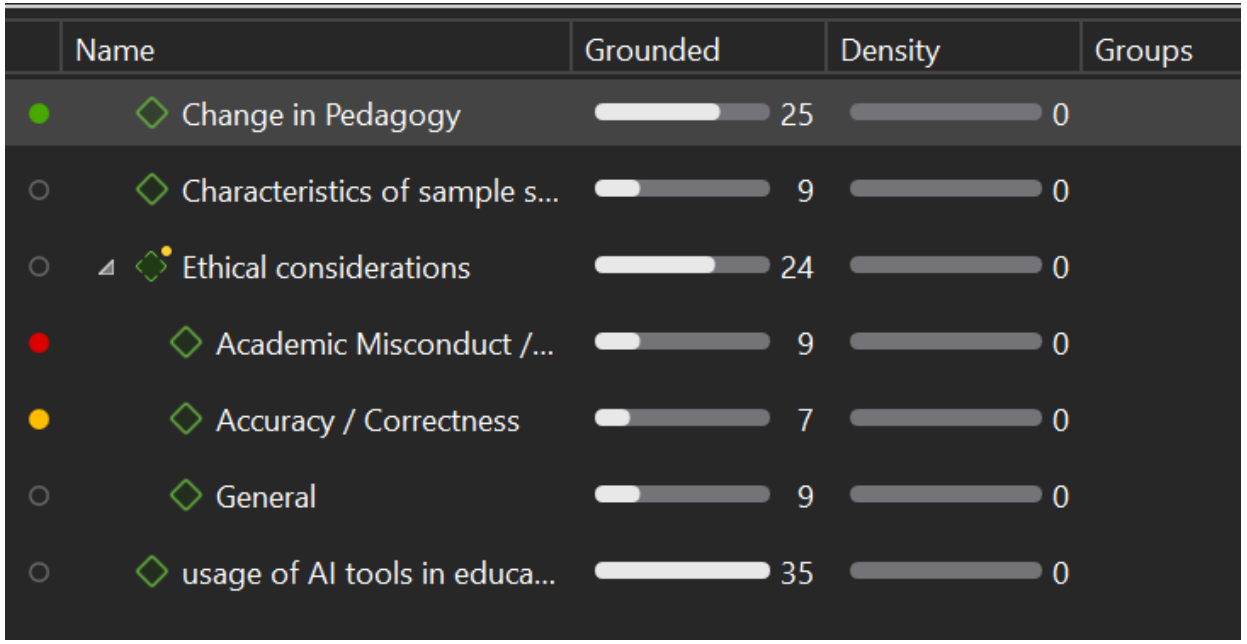
# Record Keeping

- Sorting by documents and document groups
- Part of good research is good file management
- Groupings are available for documents, codes, memos and networks
  - A file / code can also be allotted into multiple groups



# What is coding

- Coding is the process of labelling segments of data that describe what the segment is about
- It is intended to distill data and allows us to sort the data, as well as provide an analytic handle too make comparisons with other segments of data.
- At its most basic, a code is a keyword that you use to group text between and within a document. Codes may also have subcodes if your parent code is general..



# Why code data

- **Increase validity:** Qualitative coding provides organization and structure to data so that you can examine it in a systematic way to increase the validity of your analysis.
- **Decrease bias:** Qualitative coding enables you to be aware of potential biases in the way data is analyzed.
- **Accurately represent participants:** Qualitative coding allows you to evaluate if your analysis represents your participant base, and helps you avoid over representing one person or group of people.
- **Enable transparency:** Qualitative coding enables other researchers to methodically and systematically review your analysis.

# Coding Styles

## Open Codes

**Open Coding** involves identifying themes, concepts, or categories directly from the data without preconceived notions or a predefined coding scheme. It's an inductive approach where the researcher lets the data speak for itself.

- **Process**: During open coding, researchers break down the data into discrete parts and closely examine them. They identify, label, and categorize phenomena found in the data.
- **Flexibility**: This approach is flexible and allows for the discovery of new concepts and categories that may not have been anticipated.
- **Exploration**: Open coding is particularly useful in exploratory research where the goal is to uncover new insights and develop a deeper understanding of the data.

## Closed Codes

**Closed Coding**, also known as **a priori coding** or **deductive coding**, involves using a predefined set of codes derived from existing theories, literature, or research questions.

- **Process**: In closed coding, researchers apply these predefined codes to the data, categorizing segments of the data according to the predetermined themes or concepts.
- **Structure**: This approach provides a structured framework for analysis, ensuring that the coding is aligned with existing knowledge or specific research objectives.
- **Consistency**: Closed coding can enhance consistency and comparability across different datasets or studies, as the same coding scheme is applied systematically.

# Which coding style to use

- Inductive coding / ground up coding
  - Inductive coding is a ground-up approach where you derive your codes from the data. You don't start with preconceived notions of what the codes should be, but allow the narrative or theory to emerge from the raw data itself. This is great for exploratory research or times when you want to come up with a new theories, ideas or concepts.

- Deductive coding / top down coding
  - Deductive coding is a top down approach where you start by developing a codebook with your initial set of codes. This set could be based on your research questions or an existing research framework or theory. You then read through the data and assign excerpts to codes. At the end of your analysis, your codes should still closely resemble the codebook that you started off with. This is good when you have a pre-determined structure for how you need your final findings to be. For example, program evaluation studies may utilize a deductive coding approach.

# Which to use

- **When to Use Inductive Coding**

1. **Exploratory Research**: If your research is exploratory and you aim to discover new insights or develop new theories, inductive coding is appropriate. This approach allows the data to guide the identification of themes and patterns.
2. **Limited Existing Research**: When there is little existing literature or theoretical framework on your topic, inductive coding is useful as it helps generate new concepts and understandings directly from the data.
3. **Complex or Novel Phenomena**: For studying complex or novel phenomena where predefined codes might miss important aspects, inductive coding provides flexibility to capture the richness of the data.
4. **Grounded Theory Approach**: If you are using a grounded theory methodology, inductive coding is essential, as it aligns with the goal of building theory grounded in the data.

# Which to use

- **When to Use Deductive Coding**

1. **Theory Testing**: If your research aims to test existing theories or hypotheses, deductive coding is suitable. Predefined codes based on theoretical frameworks help in systematically applying and testing these concepts.
2. **Well-Established Research Area**: When there is substantial existing research and well-established theories on your topic, deductive coding ensures that your analysis aligns with and builds upon existing knowledge.
3. **Specific Research Questions**: If your study is guided by specific research questions or hypotheses, deductive coding helps in focusing the analysis on relevant themes and categories derived from these questions.
4. **Consistency Across Studies**: For studies that require consistency and comparability across different datasets or over time, deductive coding provides a structured approach that can be replicated.

# How to decide which to use

- Consider creating a decision matrix based on the following questions:

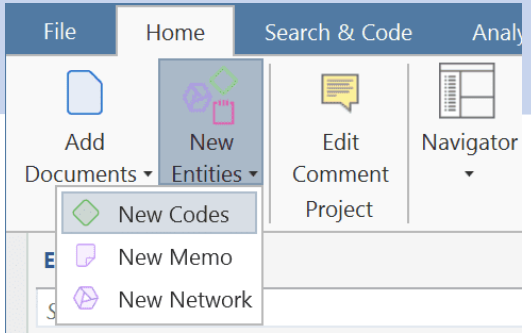
1. **What is the primary goal of your research?**
    1. Exploration and theory development → Inductive
    2. Theory testing and validation → Deductive
2. **How much existing knowledge or literature is available on your topic?**
    1. Limited or emerging area → Inductive
    2. Well-researched area → Deductive
3. **What are the nature and complexity of your data?**
    1. Complex, rich data needing flexibility → Inductive
    2. Data suited for predefined categories → Deductive

- By considering these factors, you can make a more informed decision on whether to use inductive or deductive coding for your qualitative research.

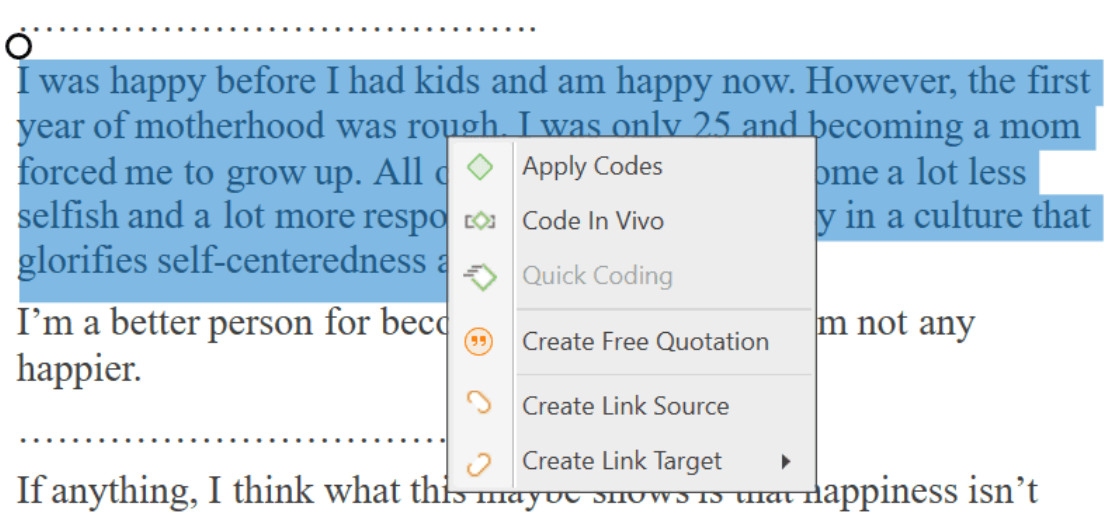
# Or use a combination

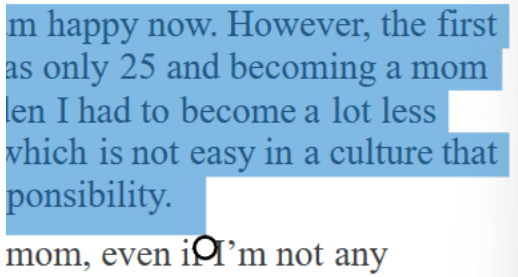
- In practice, many researchers use a combination of both approaches:
- **Start with Inductive Coding**: Begin with inductive coding to explore the data and identify emergent themes. This initial phase helps you understand the data without bias.
- **Apply Deductive Coding**: Once initial themes are identified, apply deductive coding to test these themes against existing theories or frameworks. This can refine your analysis and ensure that it is grounded in established knowledge.
- **Iterative Process**: Coding can be an iterative process where you move between inductive and deductive approaches. For instance, you might revise your codes and categories based on new data insights, then reapply them to ensure consistency.

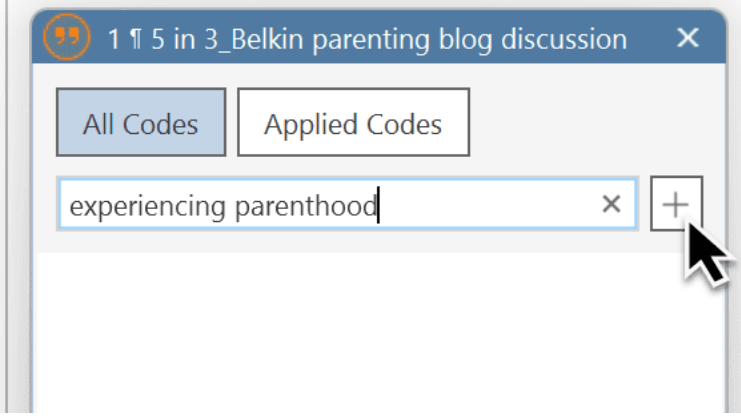
# Creating new codes




- You can create codes that have not (yet) been used for coding.
- If you already have a list of codes, possibly including code descriptions and groupings elsewhere, you can import a list of codes.
- To code a document - Right-click and select Apply Codes and then enter the definition

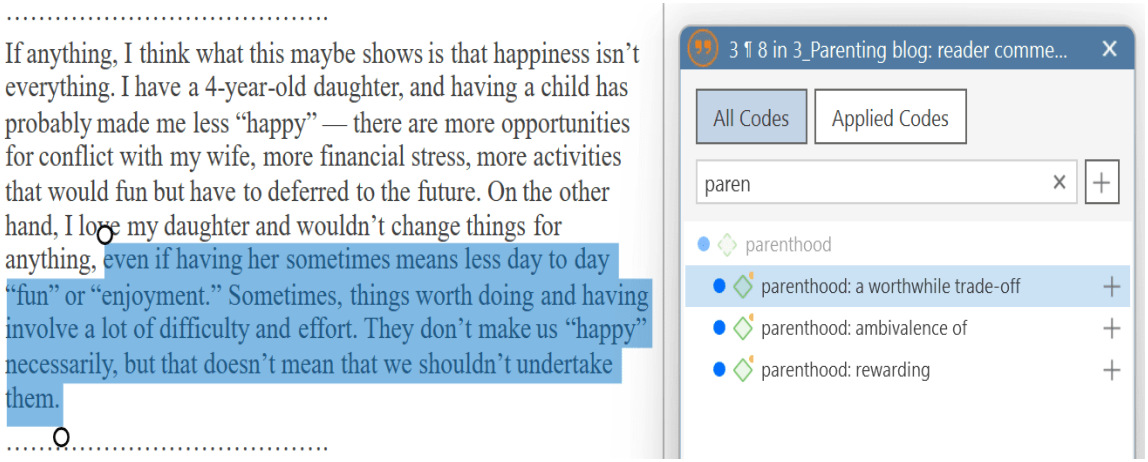


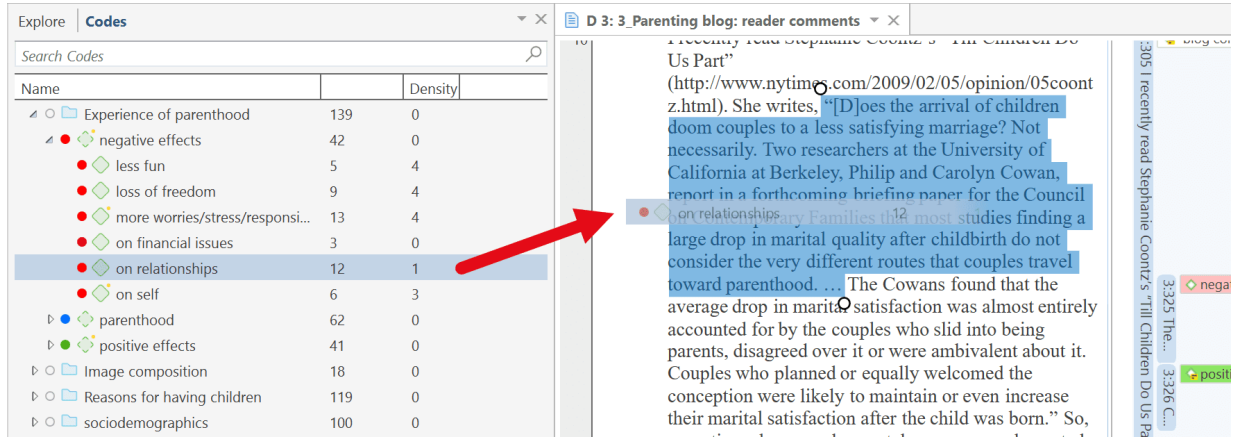



# Applying existing codes

- Highlight a data segment, right-click and select Apply Codes, or simply double-click on the quotation.

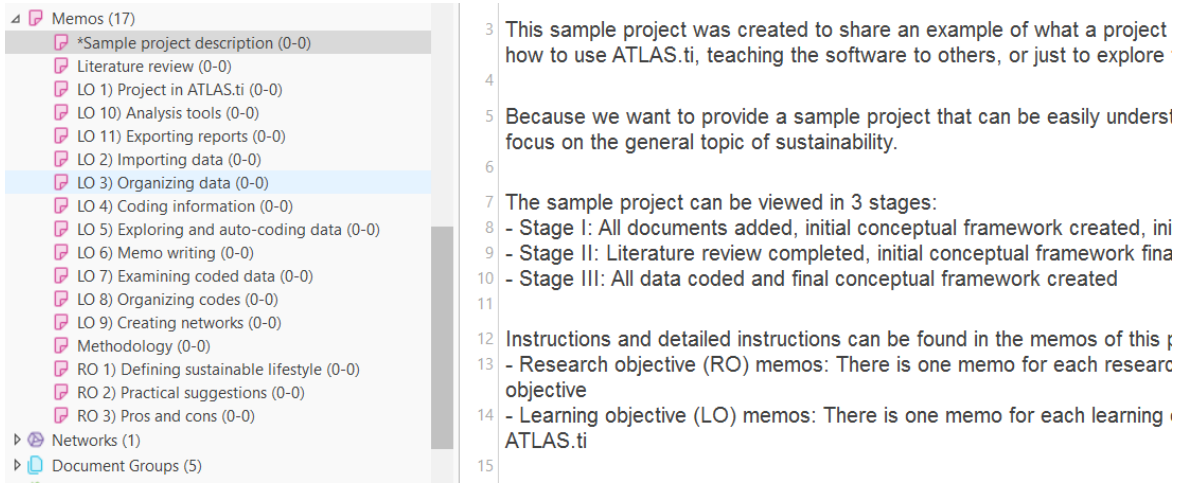



- Alternatively, open the Code Browser, highlight a data segment, select one or more codes from the code list and drag the code onto the highlighted data segment




- See more here: https://doc.atlasti.com/ManualWin.v22/Codes/CodingData.html

# Memos




- Memos can contain a **project description**
- Memos can be useful as they can keep in one place the following documents related to your research project
- The **project description**
- The **research questions** in a memo
- The **research diary**
- The **to do list**
- Randoms thoughts or observations from the relevant **literature**
- Also possible to write up your **analysis** in a memo, though depending on how fancy you want it to be, it might not be recommended

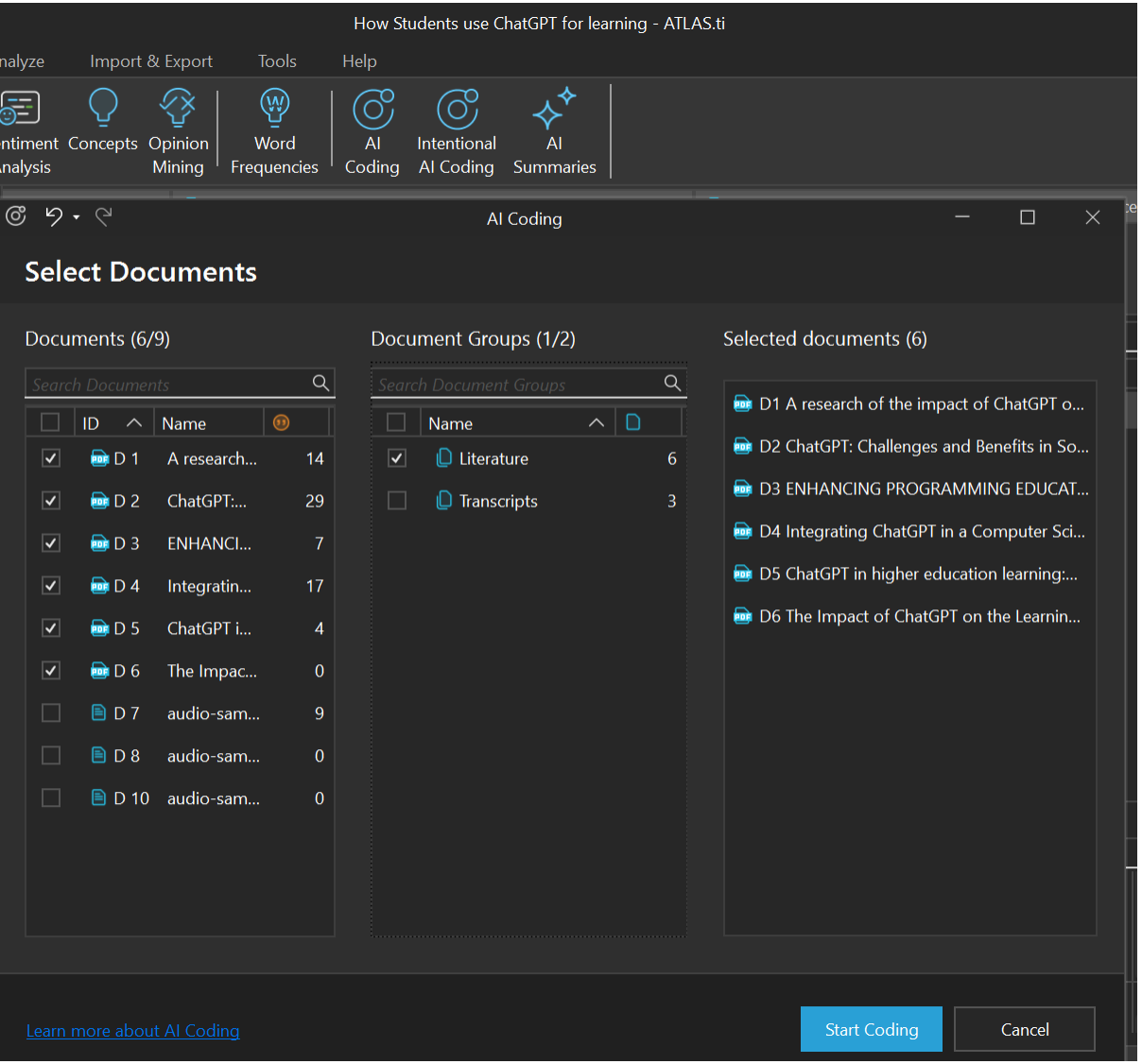
# Search & Code

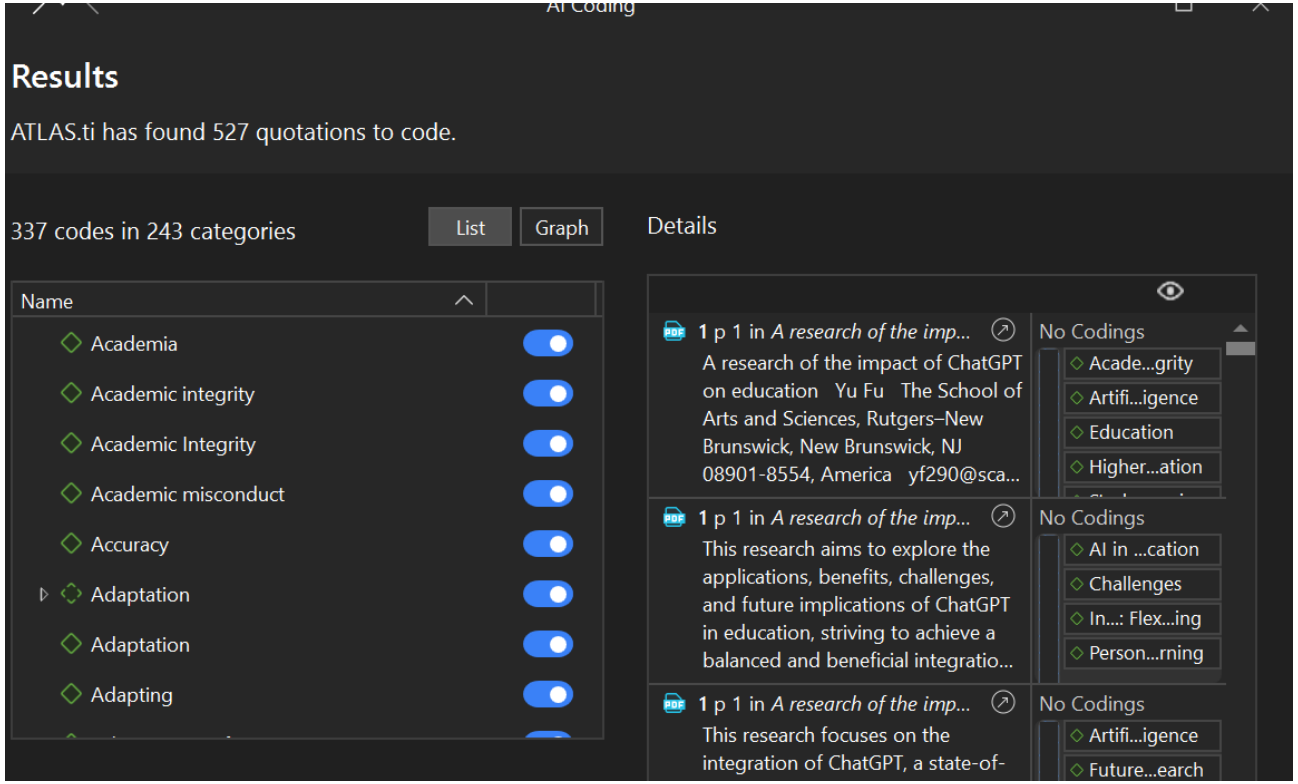
- Text Search
- AI Summaries
- AI coding & Intentional AI coding

# Tips for manual coding

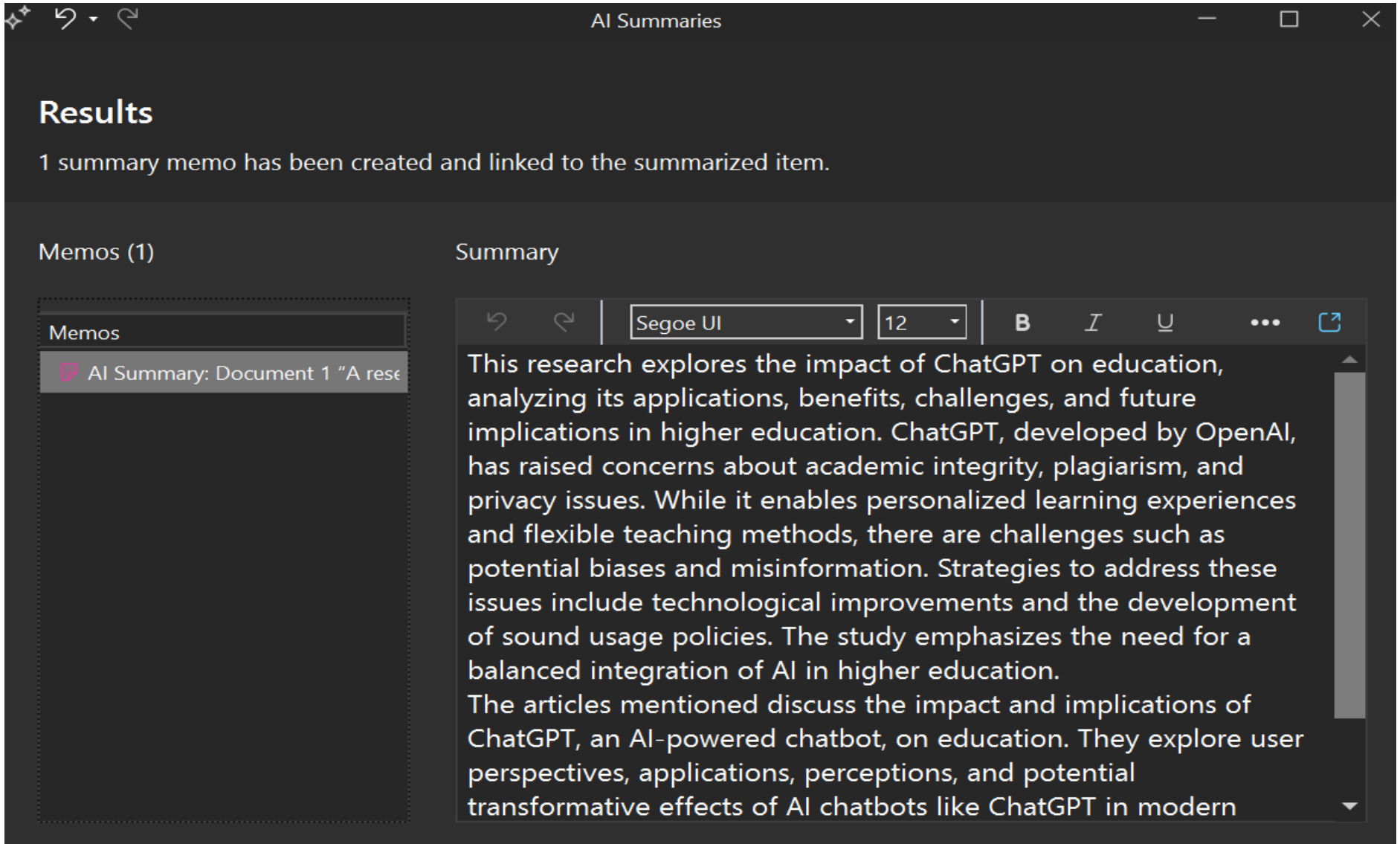
- Consider if you are going to use a **predefined set of codes** (deductive coding) or **read the material before creating the codes** (inductive coding)
  - If deductive – be mindful of bias as you might be set on proving your own hypothesis
  - If inductive – be mindful of the time required and that it is an iterative process
- Next, consider if your codes are all to have the **same level of specificity and importance** or that they have a **particular hierarchy** (i.e. some codes come within other codes)
- Keep a codebook and track the following:
  - The label used for each code
  - A description of the concept or theme the code refers to
  - Who originally coded it
  - The date that it was originally coded or updated
  - Any notes on how the code relates to other codes in your analysis
- Ensure that each code **serves a different purpose**

# AI Coding


How Students use ChatGPT for learning - ATLAS.ti
Import & Export
Tools
Help
Concepts
Opinion Mining
Word Frequencies
AI Coding
Intentional AI Coding
AI Summaries
AI Coding
Select Documents
Documents (6/9)
Search Documents
ID
Name
D 1 A research... 14
D 2 ChatGPT:... 29
D 3 ENHANCI... 7
D 4 Integratin... 17
D 5 ChatGPT i... 4
D 6 The Impac... 0
D 7 audio-sam... 9
D 8 audio-sam... 0
D 10 audio-sam... 0
Document Groups (1/2)
Search Document Groups
Name
Literature 6
Transcripts 3
Selected documents (6)
D1 A research of the impact of ChatGPT o...
D2 ChatGPT: Challenges and Benefits in So...
D3 ENHANCING PROGRAMMING EDUCAT...
D4 Integrating ChatGPT in a Computer Sci...
D5 ChatGPT in higher education learning:...
D6 The Impact of ChatGPT on the Learnin...
Learn more about AI Coding
Start Coding
Cancel

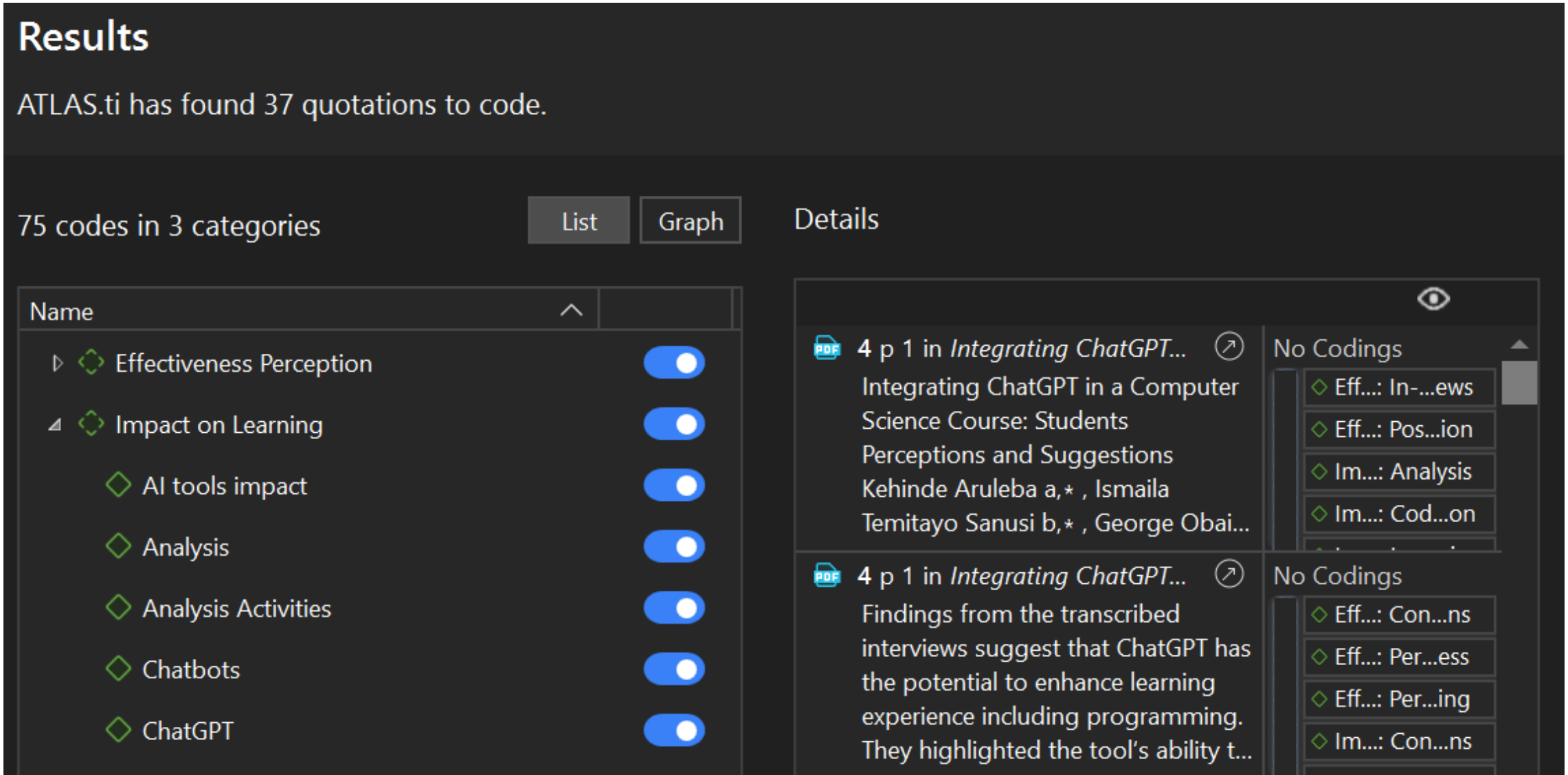

AI Coding
Results
ATLAS.ti has found 527 quotations to code.
337 codes in 243 categories
List
Graph
Details
Name
Academia
Academic integrity
Academic Integrity
Academic misconduct
Accuracy
Adaptation
Adaptation
Adapting
1 p 1 in A research of the imp...
A research of the impact of ChatGPT on education Yu Fu The School of Arts and Sciences, Rutgers–New Brunswick, New Brunswick, NJ 08901-8554, America yf290@sca...
No Codings
Acade...grity
Artifi...igence
Education
Higher...ation
1 p 1 in A research of the imp...
This research aims to explore the applications, benefits, challenges, and future implications of ChatGPT in education, striving to achieve a balanced and beneficial integratio...
No Codings
AI in ...cation
Challenges
In...: Flex...ing
Person...rning
1 p 1 in A research of the imp...
This research focuses on the integration of ChatGPT, a state-of-
No Codings
Artifi...igence
Future...earch

# AI Summary


AI Summaries
Results
1 summary memo has been created and linked to the summarized item.
Memos (1)
Summary
Memos
AI Summary: Document 1 "A rese
Segoe UI
12
This research explores the impact of ChatGPT on education, analyzing its applications, benefits, challenges, and future implications in higher education. ChatGPT, developed by OpenAI, has raised concerns about academic integrity, plagiarism, and privacy issues. While it enables personalized learning experiences and flexible teaching methods, there are challenges such as potential biases and misinformation. Strategies to address these issues include technological improvements and the development of sound usage policies. The study emphasizes the need for a balanced integration of AI in higher education.
The articles mentioned discuss the impact and implications of ChatGPT, an AI-powered chatbot, on education. They explore user perspectives, applications, perceptions, and potential transformative effects of AI chatbots like ChatGPT in modern

SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Intentional AI Coding (on literature)



# Intentional AI Coding (on transcript)

**Research Objectives:**

- To understand how novice learners are utilizing AI
- To explore perspectives from novice learners on in
- To identify potential benefits, challenges, and optii

**Research Questions:**

What are the current experiences and practices of nc

What specific coding tasks or activities do they use

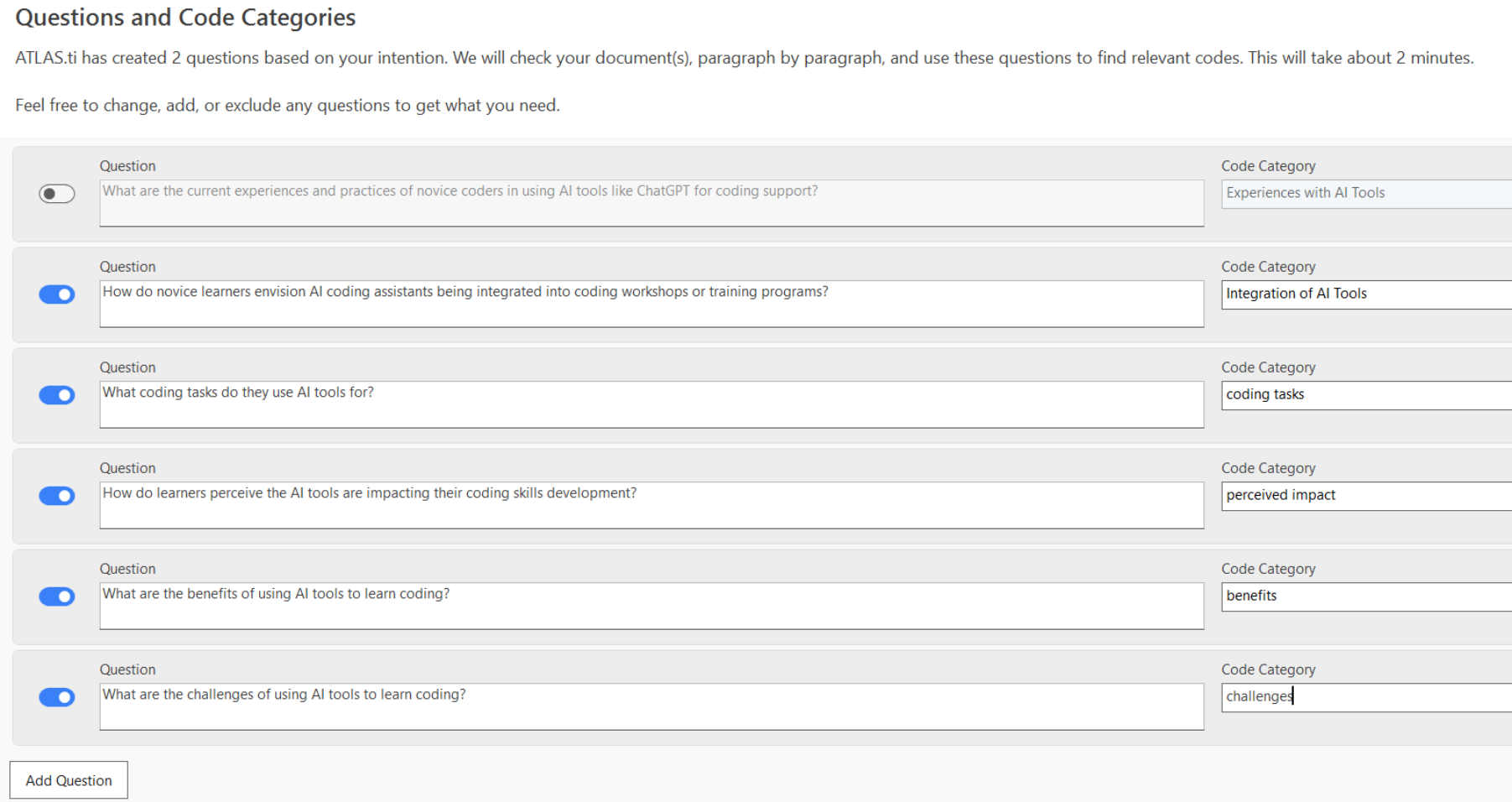
How do they perceive the tools are impacting their

What benefits or challenges have they encountered

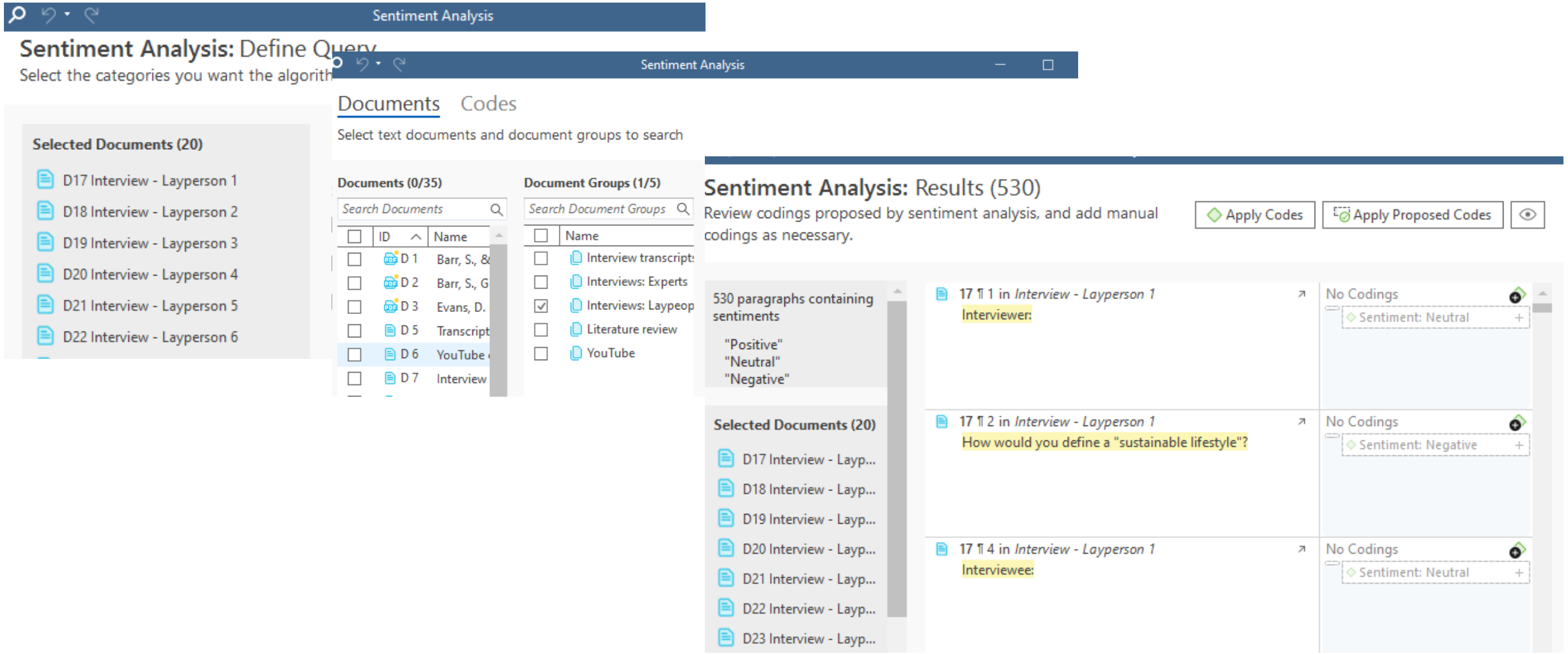
How do novice learners envision AI coding assistants

**Based on the above, please create codes into the**

- coding tasks they use AI tools for
- benefits of using AI tools for coding
- challenges of using AI tools for coding
- experience of using AI tools for coding
- integrating AI tools into workshop



# Sentiment Analysis



SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Analyzing data in ATLAS.ti

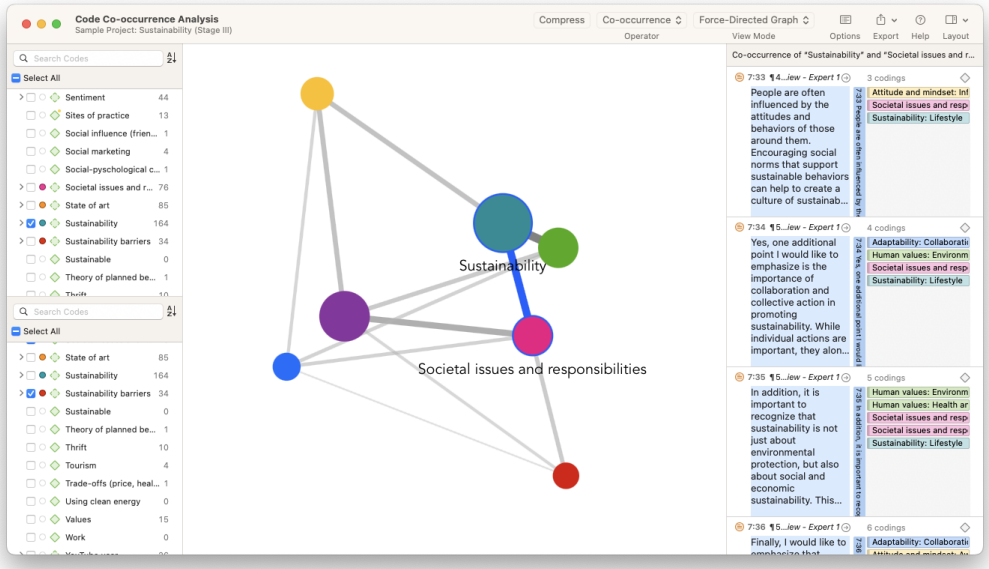
Uncover insights from your coded data

# How do I explain my analysis to others?

Quantitative vs. Qualitative:

numerical, statistical and graphical vs. textual, unstructured and interpretive

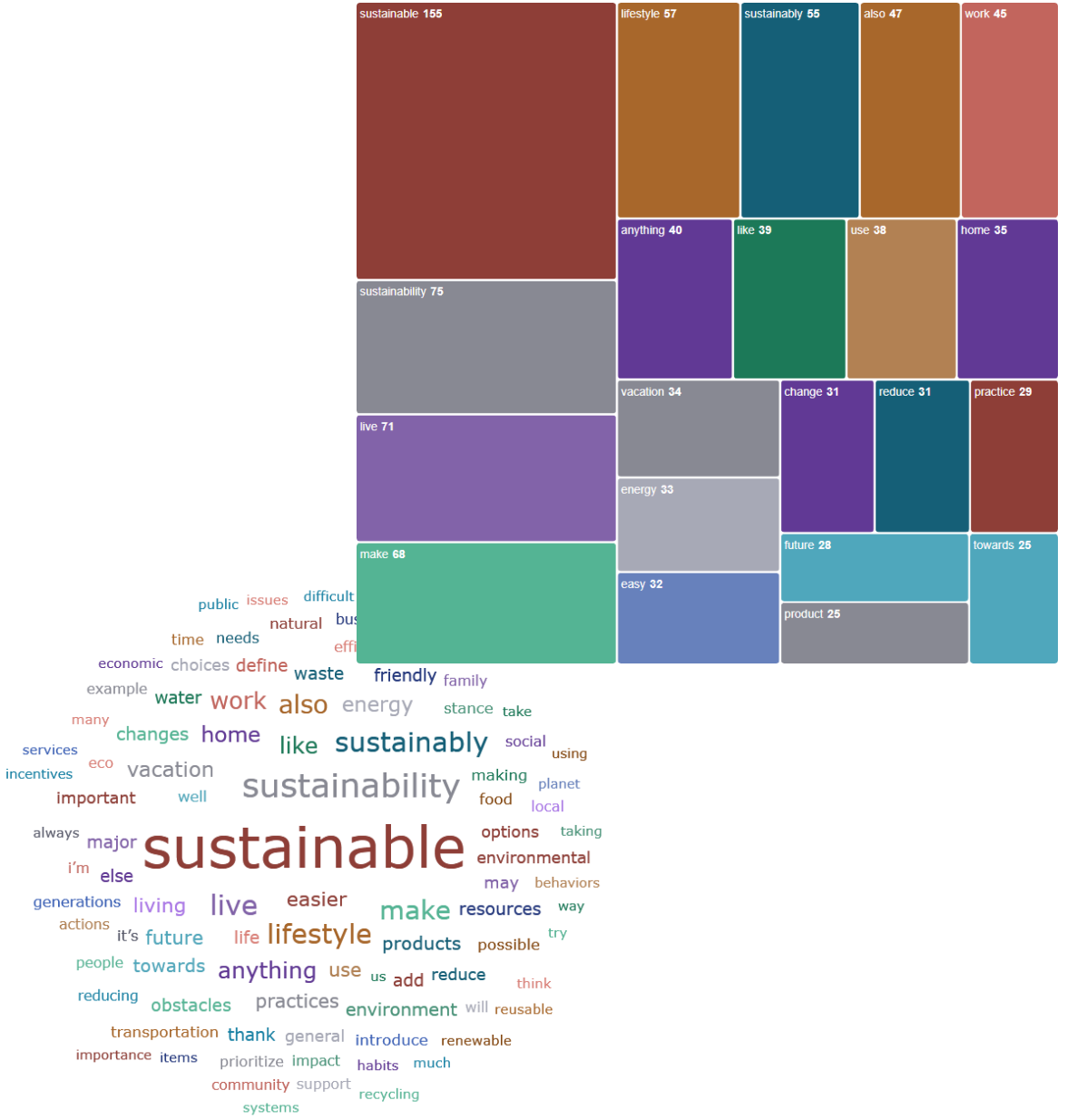
- How to present qualitative data in a meaningful and accessible way?
- **Visualizations** make complex research more comprehensible and bring clarity to key insights
- Qualitative data visualization requires that codes must adequately represent collected data



# Visualizations in ATLAS.ti

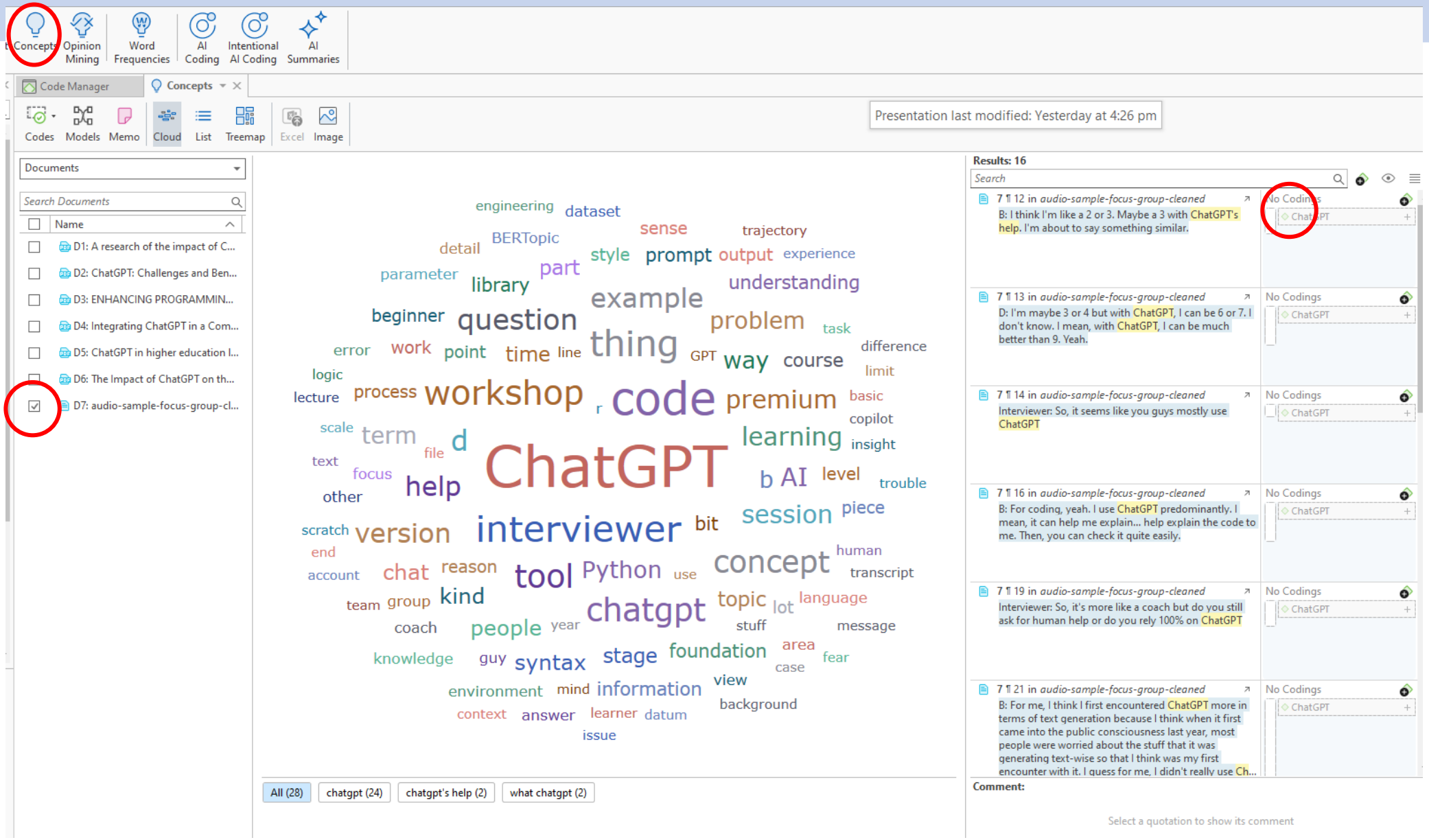
## Word Frequencies

- Shows which words appear most often in your data
  - Word Cloud
  - TreeMap
- How-to
  - `Search & Code` → `Word Frequencies`
  - Select documents/groups
  - Adjust parameters from menu options



SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Concepts



# What's the difference between concepts and word frequency?

## Concepts

- are identified using machine learning models
- significant nouns
- can be used to identify and apply codes

## Word Frequencies

- simple counts
- more flexible in configuration, e.g.
  - Edit the stop list
  - Select part of speech
  - Adjust threshold for inclusion

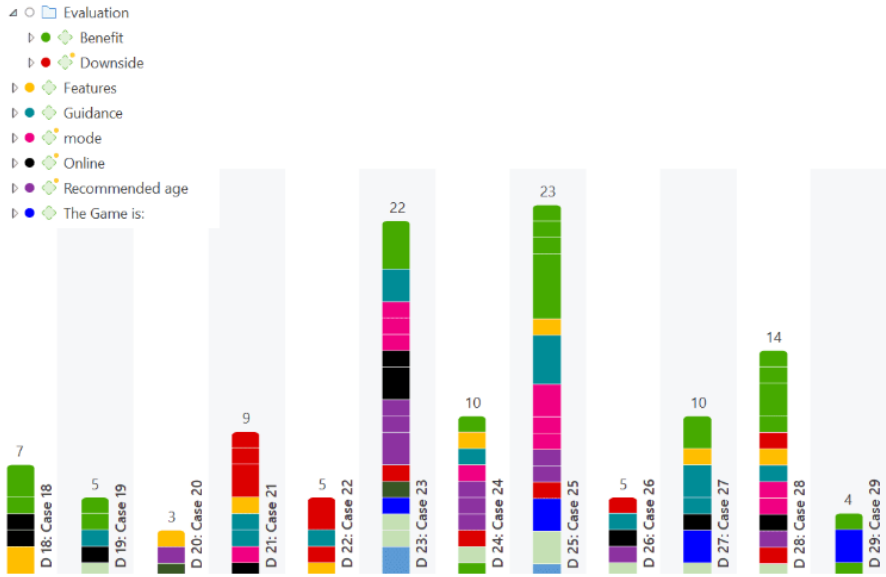
# Visualizations in ATLAS.ti

## Code Distribution across Documents

- Open `Document Manager`
- Select documents/document groups
- See distribution of codes across selected documents


## Code Distribution by Document

- Open `Code Manager`
- Select code/code groups
- See coding patterns for each document




* Check under *View* → *Diagram if you don't* see this visualization

# Co-occurrence analysis

- Relationship between codes – whether they 'co-occur' or are close to each other
- **Co-occurrence**: two or more codes occur together in the same data context
  - **Proximity**: closeness or distance between the codes, e.g. "anxiety" often appears "deadlines"
  - **Embeddedness**: one coded quotation is contained within another, indicating potential hierarchical relationships
  - The co-occur operator is a combination of WITHIN, ENCLOSES, OVERLAPS, OVERLAPPD BY and AND
- **Code co-occurrence coefficient** – strength of relation between the two codes

  **c** = n12 / (n1+n2 – n12) where n12 = number of cooccurrences of code n1 and n2

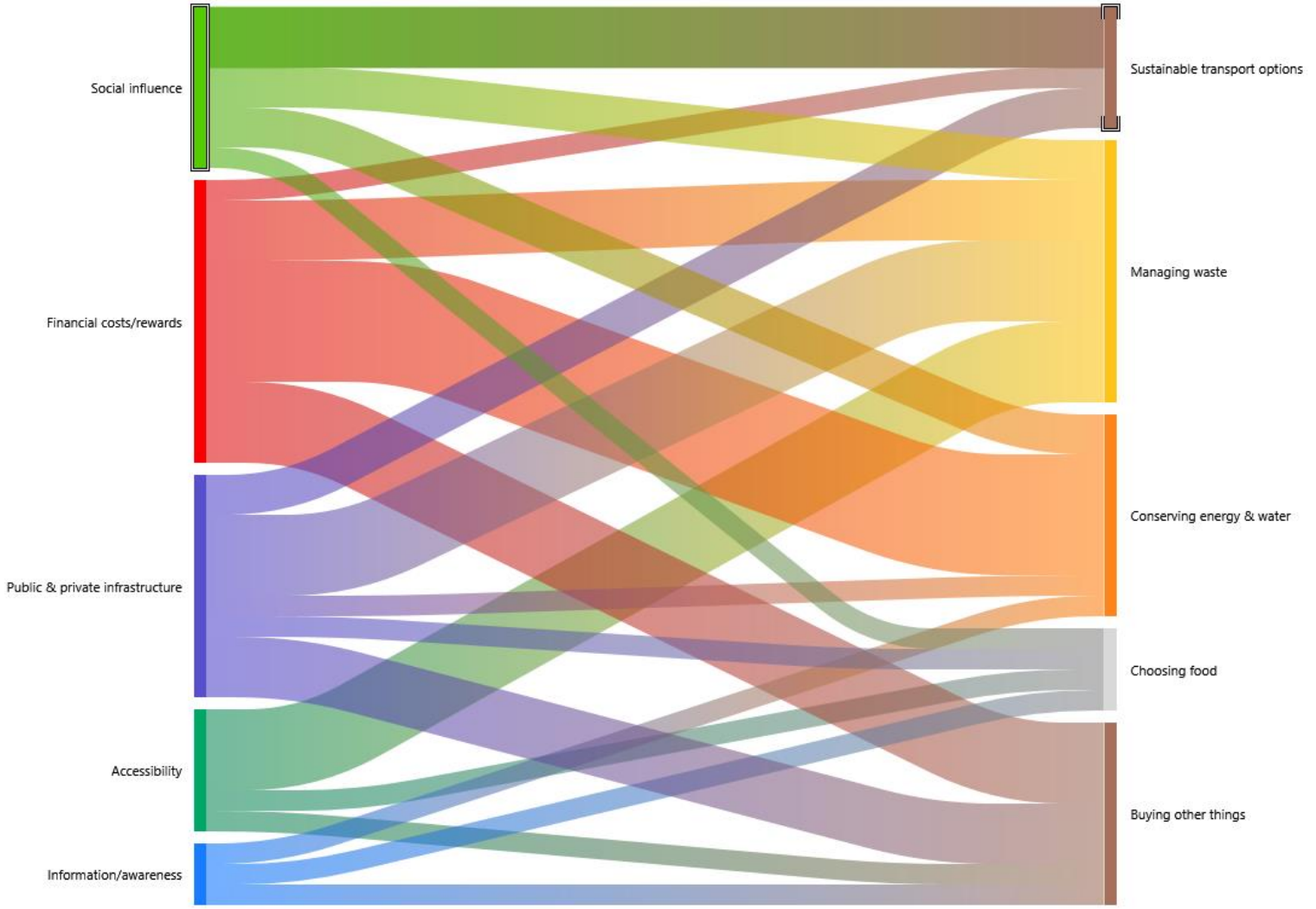
  0 – codes do not co-occur

  1 – codes co-occur at all times

  ***Note***: distortion by frequencies that differ by too much

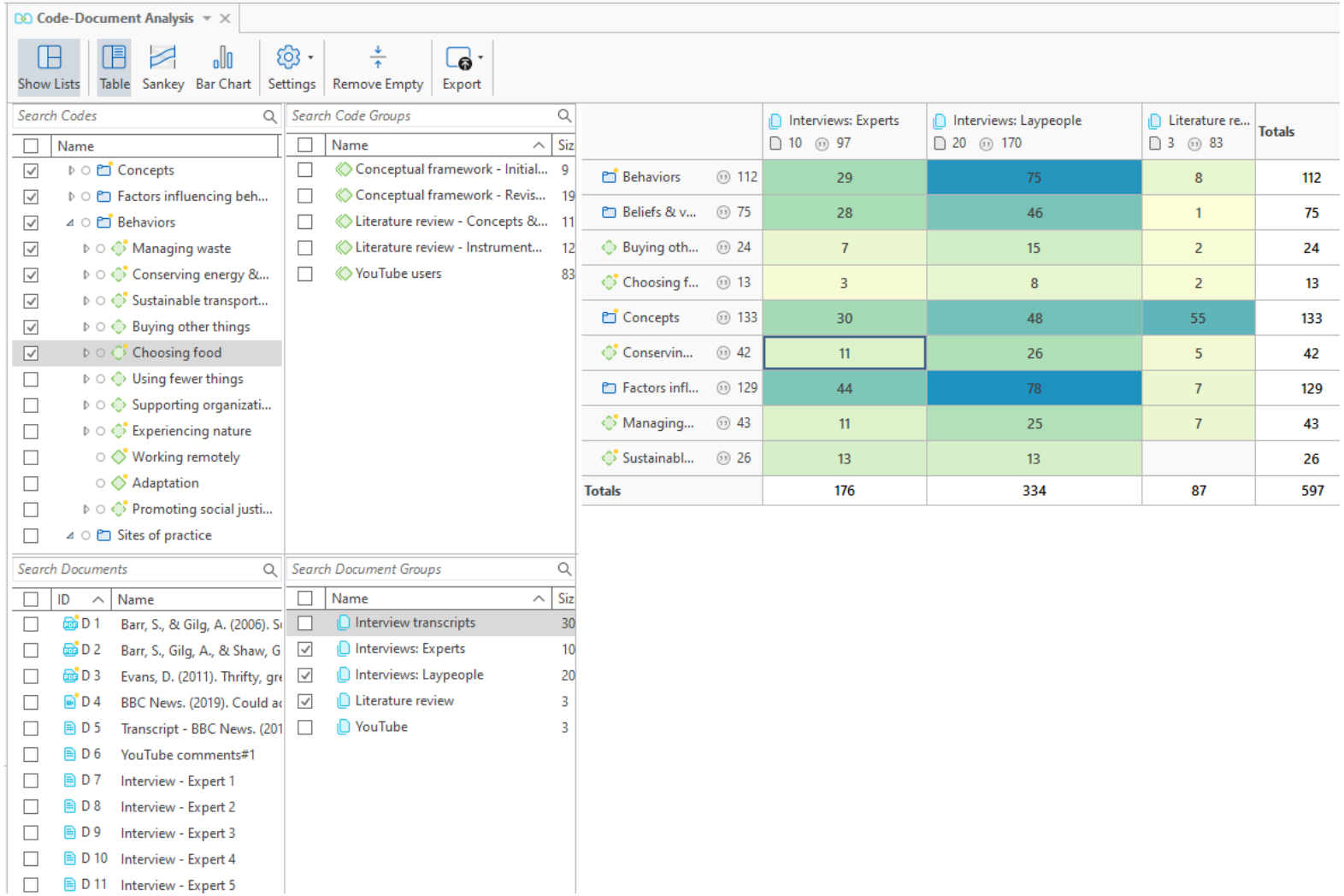
# Example: Sankey Diagram

- Available visualizations in ATLAS.ti
  - Table
  - Bar chart
  - Sankey diagram



# Code-Document Analysis

- Relate codes/code groups and documents/document groups together
  - **Rows**: select codes/code groups
  - **Columns**: select document/document groups
  - **Normalize**: Normalized frequency by no. of quotations when documents groups are of unequal size and the absolute frequencies may be misleading
  - **Absolute, row-relative and column-relative frequencies**



SMU
SINGAPORE MANAGEMENT UNIVERSITY

# Project Management & Teamwork

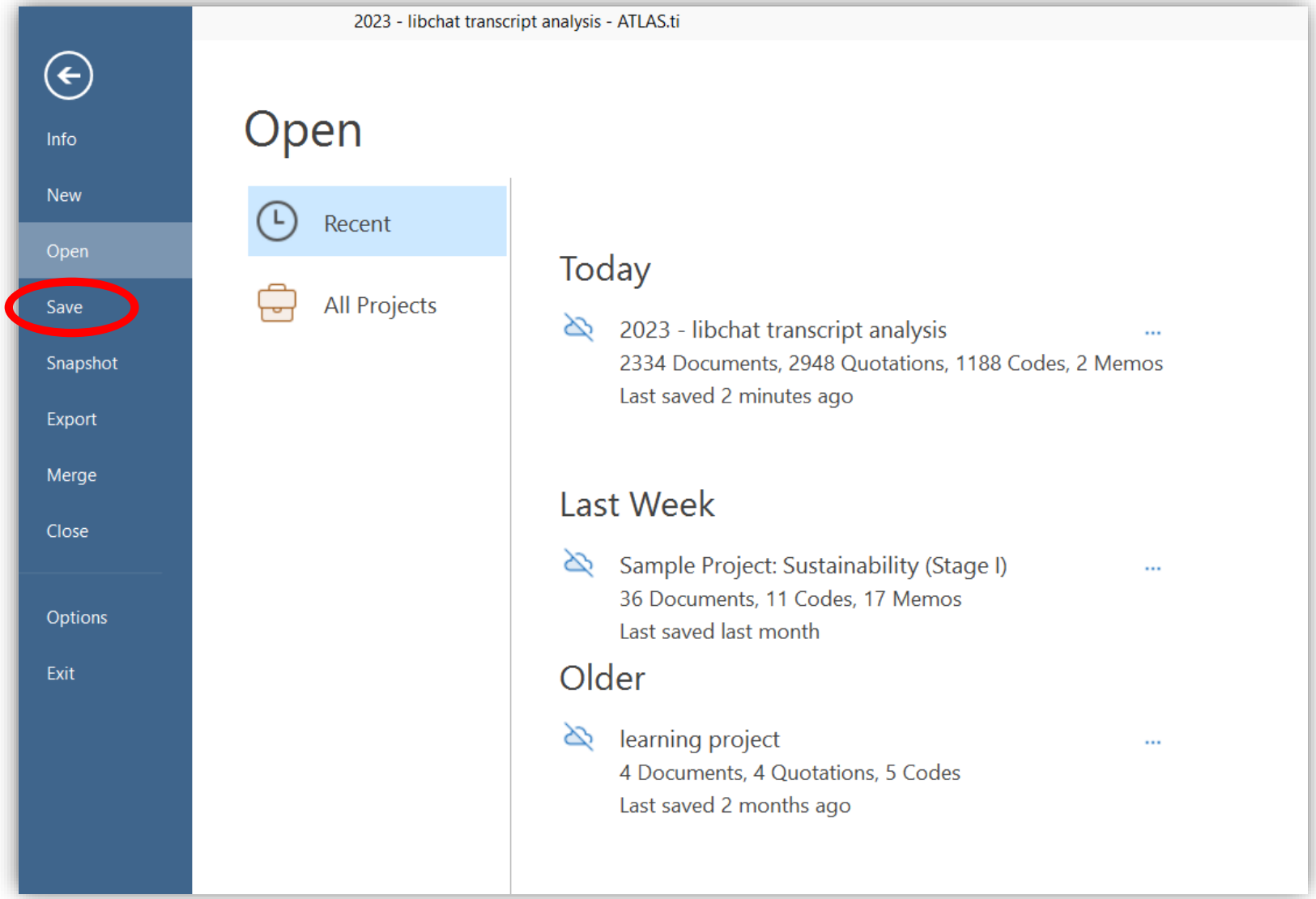
- Save, import & export projects
- How to work in teams

# How to save your project?

Click on `File` → `Save`

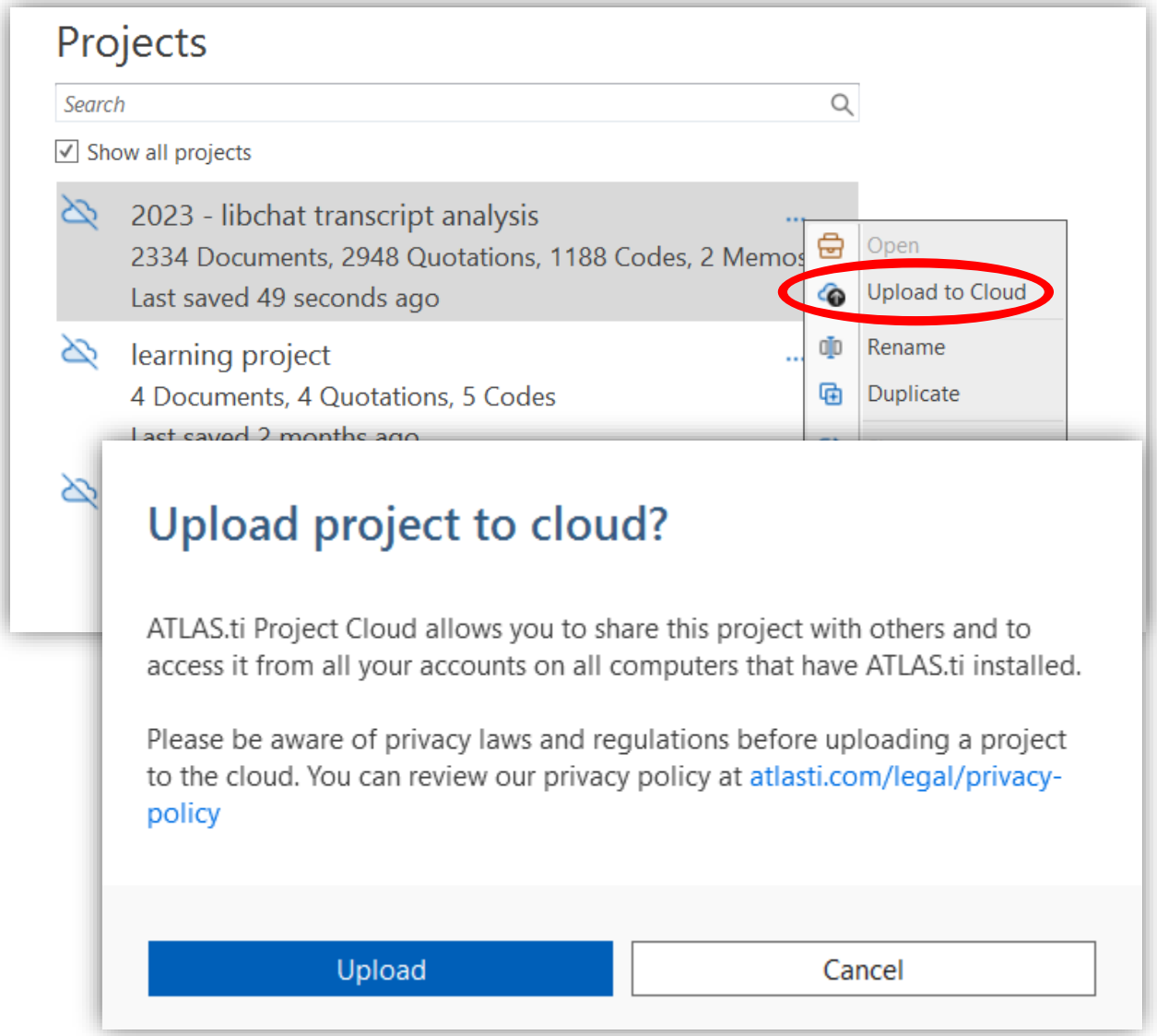
"Project Save Successful"

Save each time after using ATLAS.ti to prevent data loss!



# How to upload your project to the cloud?

- Your project will only be saved locally (on your computer) if you do not upload it to the cloud
- Projects on the cloud can be accessed from any computer
- 100MB free cloud storage space
- Be careful if you are handling personal data! (check your IRB requirement)



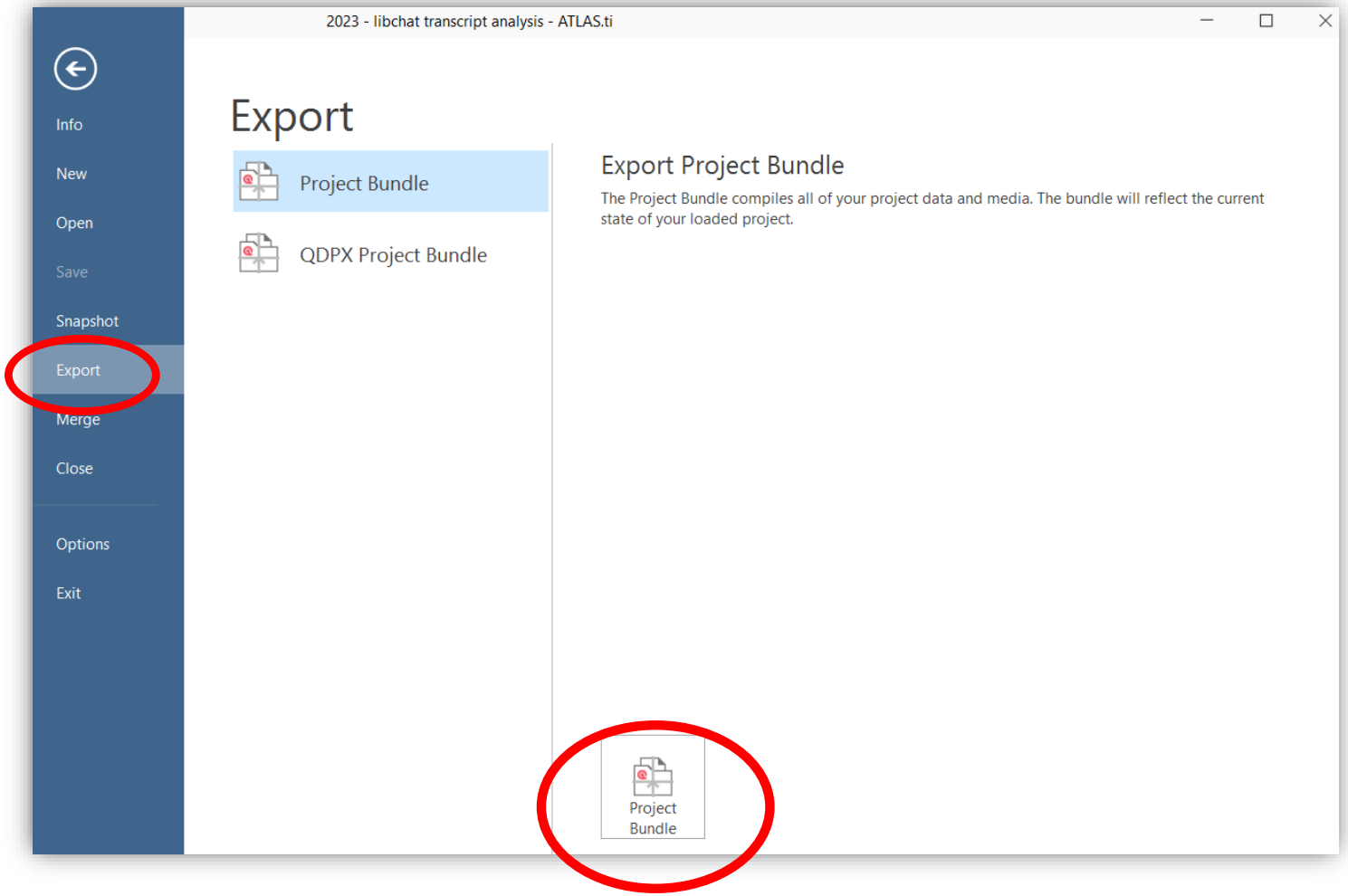
# How to export your project?

## Project Bundle

- Export as *.atlproj23* file which can be imported to another instance of ATLAS.ti software

## QDPX Project Bundle

- Export as *.qdpx* file, an XML-based structured data format
- For long-term storage
- Product-independent archival of qualitative research projects



# Team tools

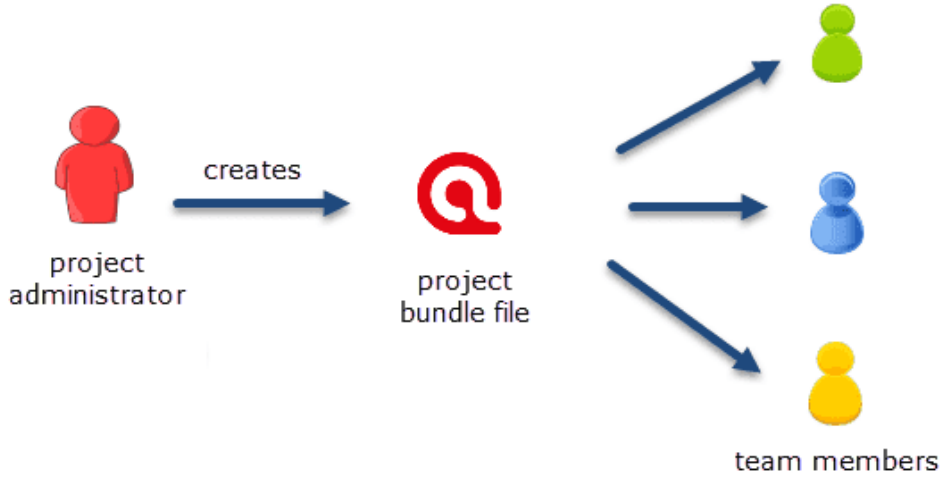
## ATLAS.ti Web version

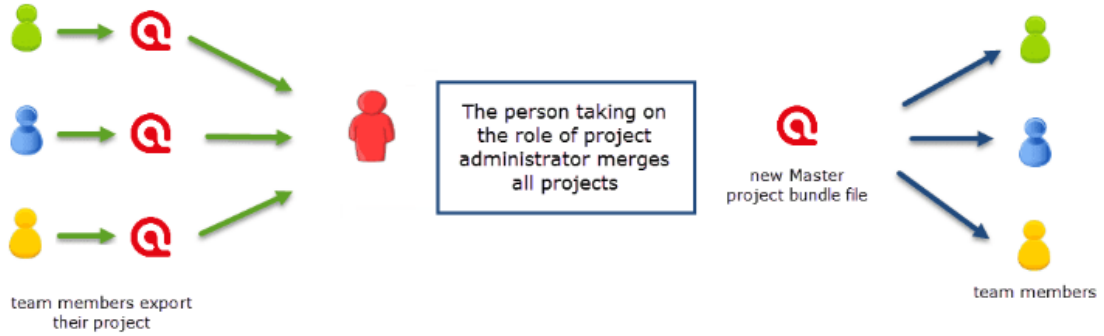
- Collaborate in real time!
- Limited features compared to the desktop version (see here for feature comparison)
- Good for early stages of your project, can be later imported into ATLAS.ti desktop

# Team tools

- ATLAS.ti Web version
- **Project Merge Method**– you and your collaborators work asynchronously and **merge** to combine your individual work. See more details here

1. Project admin sets up a Master project and add documents
2. Project admin exports the project and create a project bundle file
3. Each member import the project bundle
4. Each member codes part of the data, working on his/her own project copy
5. After individual work is done, each member sends their copy back to the project admin who will **merge** them into a new Master file



# Team tools (cont'd)

- **ATLAS.ti Web version** (for early stages of your project)
- **Project Merge Method**– you and your collaborators work asynchronously and **merge** to combine your individual work
- **Inter-Coder Agreement (ICA)**
  - Assess the agreement of how multiple coders code a given body of data
  - To ensure reliability of the coding process, and that different coders interpret and apply codes in a consistent manner
  - Built based on Prof. Klaus Krippendorff's work Content analysis: An introduction to its methodology
  - Available tests: *simple percent agreement*, *Holsti Index* and *Krippendorff alpha coefficients*
  - Find out more about ICA on the manual

# References, Learning Resources and Guides

- ATLAS.ti User Manual for Windows and Mac
- Official video tutorials
- The Ultimate Guide to Qualitative Research
  - Part 1: The Basics
  - Part 2: Handling Qualitative Data
  - Part 3: Presenting Qualitative Data
- ATLAS.ti AI Lab – AI tools

Tuesday, 28 May - Thursday, 30 May 2024

# AI for Research week

# AI for Research Week 2024

## Programme & Registration

smu.sg/aiweek2024

# Thank You!

## Questions?

Email us at library@smu.edu.sg

# Readings to explore

- https://methods-sagepub-com.libproxy.smu.edu.sg/search/results?methods=[%22ATLAS.ti%22]
- Bingham, A. J. (2023). From Data Management to Actionable Findings: A Five-Phase Process of Qualitative Data Analysis. *International Journal of Qualitative Methods*, *22*. https://doi.org/10.1177/16094069231183620
- Kalpokas, N., & Radivojevic, I. (2022). Bridging the Gap Between Methodology and Qualitative Data Analysis Software: A Practical Guide for Educators and Qualitative Researchers. *Sociological Research Online, 27*(2), 313-341. https://doi-org.libproxy.smu.edu.sg/10.1177/1360780421100357 9