



Analysis: How will we make sense of what people said?

If we know that collecting, combining, and interpreting people's stories inevitably changes their meaning, how do we design a system that does this responsibly?



Connect Considering the activities from earlier in this session, what are some ideas we should consider when *analyzing conversations*?

Build See the following pages for examples of what this looks like.

Who reads the full conversations

And how much time will this take?

Naming themes

How you'll sort and make sense of this data.

AI, if any

Which step, which approved tool, who uses it.

Keeping context

How each quote stays tied to its question and context.

What doesn't fit

Where contradictions and unique stories go.

Students' check

When participants hear and challenge your themes.

Revise After creating your process, make sure it accounts for the amount of time this will take during a busy school year. Analysis tends to be the most intensive part of this work.

Translating into findings

Consider the following example.

① What students said Ninth graders, several conversations about belonging	- a “In middle school everybody knew me. Here I could go a whole day without anyone saying my name.” - b “Robotics is the only place I feel like I’m supposed to be. The teacher noticed I was good at wiring, and now I run the electrical team.” - c “I quit volleyball because I have to watch my brothers after school. It’s not that I didn’t like it.” - d “Advisory is cool. I like that my teacher did a bunch of team-building games at the beginning of the year and now I know everyone.” - e “I don’t really need to belong here. I have friends at home, and I just want to get through the school day.” - f “The upperclassmen in band kind of adopted us. They showed us where to sit at lunch.”
② Accessing this information	After transcribing your recordings, you’ll need a way to organize the data, such as: • Highlighting transcripts by hand (with a highlighter, or in Google Docs). We recommend highlighting every single point made, not just the ones that stand out. • Using AI to identify “conversation turns” (each time a new person begins to speak, or makes a separate point). • Along with either of the above, making flashcards (printed or digital) so the information is easy to sort.
③ What we notice	Start by closely reading each quote and making sure we understand what it means.
④ A possible theme	Then consider a theme that’s started to emerge across your first quotes: Being noticed. For several students, belonging was associated with an adult or an older student noticing them. We’ll sort these into a group, either physically (with flashcards) or with a note on the transcript (a “code” from an emerging codebook). For example: Being noticed, b d f. Keep going, quote by quote, until clear themes emerge. If a single theme collects a ton of quotes, break it down into subthemes (or into different themes entirely). If some ideas don’t fit anywhere, make a new code or simply label them “outliers.” We also like marking standout quotes with a star or symbol so we can highlight them later.

Especially if doing analysis with students, start simple

Our student co-analysis sessions (free at pd.humanrestorationproject.org) use a deck of 36 cat photos: students sort them into themes, re-sort them into three completely different ones, and then find subthemes. Groups decide differently what the categories should be. This helps define what a theme is, and shows that disagreeing doesn’t mean one side is wrong.

Expect your first conversation to be more critical

The first time students (or adults, for that matter) get an open mic to talk about school, they often lead with frustrations. That’s normal, and it isn’t proof that everything is wrong. Many of the concerns (school feeling repetitive, wanting more say, needing breaks) show up in schools everywhere, so they can be taken seriously without anyone blaming themselves. These are systemic concerns. As your community sees change happen because of these conversations, the conversations tend to become much more positive and proactive.

The MIRACLE framework

The MIRACLE framework is a checklist from Ahtisham Younas and colleagues (2023), a team of nursing and education researchers at universities in Canada, Spain, Colombia, Pakistan, and the United Kingdom. They wrote it to help people understand what high-quality “thick data” looks like. Below, we’ve crosswalked their concepts with the algorithmic biases of AI described in Holmes and Porayska-Pomsta (2022), and what to do to consider or prevent each.

MIRACLE framework	Algorithmic bias	What you can do
Meaningful: Statements capture feelings and are accurate (for example, supporting quotes and storytelling).	Deployment: A mismatch between how the AI tool is designed and what it’s being used for.	All claims are backed up by ample quotes and data, and we use intentionally designed tools to support that work.
Interpretative: Underlying meanings, intents, and motives are captured (for example, supplementary research and historical context are provided).	Historic: Moral, legal, and sociocultural biases are present in AI.	Reporting clearly says why and how these interviews were conducted, and our tools are given the context for what’s being shared and why.
Relational: Recognizing how participants express themselves in their world (for example, tagging and highlighting how different demographic groups participate).	Learning: Misinformed or unverified assumptions and methods in the model.	As an optional step, capture demographic data from participants. When reading data in our tools, we’re familiar with and question any presumed ideas.
Authentic: Using participants’ genuine viewpoints with their exact experiences (for example, including native languages and slang).	Aggregation: One model treats every group the same, even when the same words mean different things to different groups.	Share context with every quote, define slang as needed, and use high-quality translation.
Contextualized: Having adequate detail about community, cultural orientations, and timeframes.	Representation: Over- or underrepresentation of some populations, and methodology biases.	Include as many community members as possible in the conversations, not just a select few.
Linked: Linking to a research question set or theoretical framework, with quotes grouped thematically.	Measurement: Variation and accuracy of labels across conversations and groups.	Have multiple people review themes and challenge findings, and make sure any AI tool has a “human in the loop” step.
Emic: Offering “micro details” about an individual’s experiences on sensitive topics by sharing “just enough,” without labeling or stereotyping.	Evaluation: The tool was tested on data that doesn’t look like the people it’s being used with.	Avoid sharing personally identifiable information, without losing sight of what makes a perspective unique.

At a glance

Start with highlights

A highlight is any statement with a clear point: “I wish we had more time to go outside.” “The teachers at this school care about me.” Mark them across every transcript before you group anything.

Decide what counts as a theme

Agree on a threshold before you start. In a recent empathy interview session, we decided a shareable theme had to show up in at least 5% of the highlights.

Say how strong the evidence is

In your reporting, consider grouping evidence into categories like “strong,” “moderate,” and “emerging,” based on the share of highlights behind each theme.

Keep the counterevidence next to the finding

If 41 conversations described school discipline as generally fair, and 37 described students with a past being judged more harshly, share both side by side. Tension like this in qualitative research lets us dig deeper and have powerful conversations.

Where AI fits, and where it doesn't

Language models predict statements from patterns in a huge amount of text. That makes them particularly useful for sorting through large amounts of qualitative information. In fact, that was a major focus of AI before modern large language models like ChatGPT and Claude. However, we have to be careful that our AI tools don't make our decisions for us.

Ask a language model to summarize the six belonging statements on page 2 and you may get something like “Students value connection, recognition from adults, and a sense of community.” It may leave out outliers, like the student who doesn't need to belong at school. Or it may sound so complicated that it's hard to make sense of.

People and models do the same things to meaning

	When people do it	When AI does it	What to check
Interpret	We tend to see things in the context of our own experiences and perspectives.	It analyzes in the context of its training data.	Given the context of the conversation, what are people actually saying?
Compress	We find patterns and distill them into easy-to-understand chunks.	It finds patterns but often overcomplicates findings, or overwhelms with more than a person can take in.	Are we missing information as we try to make sense of it?
Assume	We make assumptions, from our biases about what's working or not, based on how we interpret the findings.	It fills in “missing information” from its training data, sometimes without telling you.	How can we bring in multiple perspectives, including the participants themselves, to make sense of this data?
Frame	We present data in ways that are kind or critical toward ideas we agree or disagree with.	It's often overwhelmingly positive or negative when describing outcomes.	How can we tell what really happened while encouraging meaningful outcomes?

If you're comfortable with it, these are all reasonable uses for AI:

- Transcribing and translating
- Finding every excerpt about a topic
- Pulling quotes for a theme you've already named
- Searching the whole set for counterexamples to your theme
- A rough first report that you'll rewrite

But people should always be responsible for:

- Deciding what a student meant
- Deciding what matters, and which stories should be shared first
- Deciding whether a finding is ready to share, and with whom
- Deciding what the school will change

At Human Restoration Project, we use AI tools for qualitative analysis. But we keep people in the loop for:

- Checking representative samples for the codebook (for example, checking 10% of the “conversation turns” for accurate AI tags)
- Reading through all findings and making sure they connect back to the themes
- Making sure findings are written in an engaging, easy-to-read, and kind way
- Searching for missing information in the codebook or reports

Offloading the right part

We “offload” our thinking to tools all the time: calculators, calendars, spell check. Researchers call this “cognitive offloading” (Risko and Gilbert, 2016), and it often frees us up for better work. It becomes a problem when we stop thinking critically about our tools and what they find. In a 2025 survey of 319 knowledge workers, people who were more confident in AI reported doing less critical thinking when they used it (Lee et al.). Read through the transcripts before you read the summary, and ask: would we reach the same conclusion if we had to defend it with the students' words alone?

Making strong prompts

A quick prompt: “Identify the key themes in these transcripts.” It invites the model to average everyone into a few tidy headings and to fill gaps with what's typical.

A stronger approach works in multiple rounds, with people making the decisions in between (a “human in the loop”). In two studies of AI-assisted thematic analysis, the model suggested codes, people reviewed and debated them, and the final set of themes came out of that discussion (Dai et al., 2023; Gao et al., 2024). Dai's team found this back-and-forth matched the quality of coding done by people alone, in less time.

First, read on your own

Before anyone opens an AI tool, each person on the team reads two or three transcripts and jots down possible themes. You'll check the model against these notes, and they keep its first answer from becoming everyone's answer.

Step 1: Find possible themes (an AI prompt)

You're helping a school team analyze student empathy interviews about [your focus]. The de-identified transcripts are attached. Each conversation has an ID (C01, C02...), and each quote is numbered.

1. Go through every transcript. List each distinct point a student makes as an exact quote, with its conversation ID and quote number. Don't paraphrase, and don't skip any quotes.
2. Group the quotes into possible themes. For each theme, give a short name in students' own words, a one-sentence definition, and every quote that supports it.
3. Under each theme, list any quotes that complicate or contradict it.
4. List the quotes that don't fit any theme. Don't leave them out.
5. Count how many conversations each theme appears in, using your quote list. Don't estimate.

Keep what students said separate from your interpretation. If something isn't in the transcripts, say so.

Then, people in the loop

- **Compare.** Put the model's themes next to your own notes. Which did you both find? What did it find that you didn't, and what did you notice that it missed?
- **Check.** Search the transcripts for a few quotes under each theme to confirm they're word for word and come from the conversation it names.
- **Read the leftovers.** Look closely at the quotes that don't fit and the ones that contradict a theme, and decide whether each is an outlier or the start of a theme the model missed.
- **Revise.** Merge, split, rename, or drop themes until they match what you see. Where people disagree, go back to the quotes together. Write each theme as a name and a one-sentence definition.

Step 2: Sort every quote into your themes (an AI prompt)

Here are the themes our team agreed on, each with a name and a one-sentence definition. Sort every numbered student quote in the attached transcripts into these themes. Return a table in CSV format with these columns: Conversation ID, Quote number, Exact quote, Theme(s), Reason (one sentence). A quote can fit more than one theme; use "None" if none fits. Don't change our themes. If several quotes point to a theme we're missing, list them separately at the end.

Then check a sample in a spreadsheet, as in the steps below. If the checks turn up problems, sharpen the definitions or add the missing theme, and run Step 2 again.

Running this in a ChatGPT or Claude project

A project keeps your instructions, transcripts, and themes together, so every chat starts from the same files and rules. Whether you use NotebookLM, ChatGPT, or Claude, put rules like these in the project's instructions:

You're helping a school team read de-identified transcripts of student group conversations. Quote students exactly, and say which transcript each quote is from. Keep what students said separate from your interpretation, and label interpretations as possible readings. When students disagree, show each side. Point out unusual or one-of-a-kind comments rather than leaving them out. Don't say how many students said something unless you counted. If something isn't in the transcripts, say so.

- ① **Use the right account.** Create the project in the version your district has approved for student data (usually an education or enterprise plan), not a personal login.
- ② **Add the instructions.** Paste the rules above into the project's instructions.
- ③ **Prepare the transcripts.** One file per conversation, named by ID (C01, C02...), with names removed and each quote (each time someone speaks) on its own numbered line. Upload them to the project.
- ④ **Read, then find themes.** Each person reads a few transcripts first. Then run Step 1 in a new chat, and compare, check, and revise the themes as a team. Upload your theme list to the project as a file.
- ⑤ **Sort everything.** Run Step 2. Copy the CSV table it returns into Google Sheets or Excel.
- ⑥ **Check a sample.** Add a column with `=RAND()`, sort by it, and check at least one row in ten against the original transcript, two people if you can. Search for each quote to confirm it's word for word, and mark whether the theme fits. If more than about one in ten is wrong, sharpen the theme definitions and run Step 2 again.
- ⑦ **Count in the spreadsheet.** Use a pivot table or `COUNTIF` to see how often each theme appears, rather than trusting the model's own counts. Keep the "None" rows and the contradictions in view.
- ⑧ **Keep a record.** Save each version of your prompts and themes, so someone else could repeat the process and get to the same place.

What the checking sheet looks like

Conv.	Quote #	Exact quote	Theme(s)	Reason	Quote exact?	Theme fits?
C04	12	"The upperclassmen in band kind of adopted us."	Being noticed	Older students gave them a place	Yes	Yes
C07	31	"I quit volleyball because I have to watch my brothers."	Being noticed	About leaving an activity	Yes	No: time and family
C02	8	"I just want to get through the school day."	None	Doesn't fit a theme	Yes	Yes

Before any transcript goes into a tool

- Only de-identified transcripts, and only in a tool your district has approved for student data. A personal account is not the same as the district's version of the same product.
- Removing names may not be enough. For example, "The only junior on the robotics team" identifies someone.
- Check what your consent letter told families. If it didn't mention AI, ask before you use it.

Younas, A., Fàbregues, S., Durante, A., Escalante, E. L., Inayat, S., & Ali, P. (2023). Proposing the "MIRACLE" narrative framework for providing thick description in qualitative research. *International Journal of Qualitative Methods*, 22.

Holmes, W., & Porayska-Pomsta, K. (Eds.). (2022). *The ethics of artificial intelligence in education: Practices, challenges, and debates*. Routledge.

Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*, 20(9), 676–688.

Lee, H.-P., et al. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *CHI '25*.

Dai, S.-C., Xiong, A., & Ku, L.-W. (2023). LLM-in-the-loop: Leveraging large language model for thematic analysis. *Findings of the Association for Computational Linguistics: EMNLP 2023*, 9993–10001.

Gao, J., et al. (2024). CollabCoder: A lower-barrier, rigorous workflow for inductive collaborative qualitative analysis with large language models. *CHI '24*.

