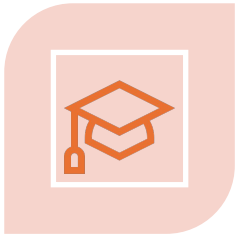
INSIGHTSTACK: From Transcripts to Insights

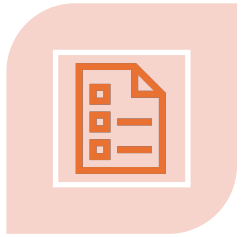
Christabel Acquaye, Jordan Hosier, Yu Zhou, Vijay K. Gurbani

Presentation Outline

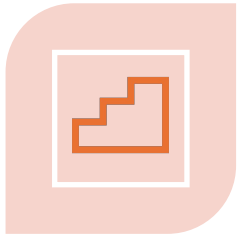


BACKGROUND

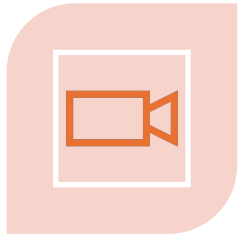
METHODS



EVALUATION

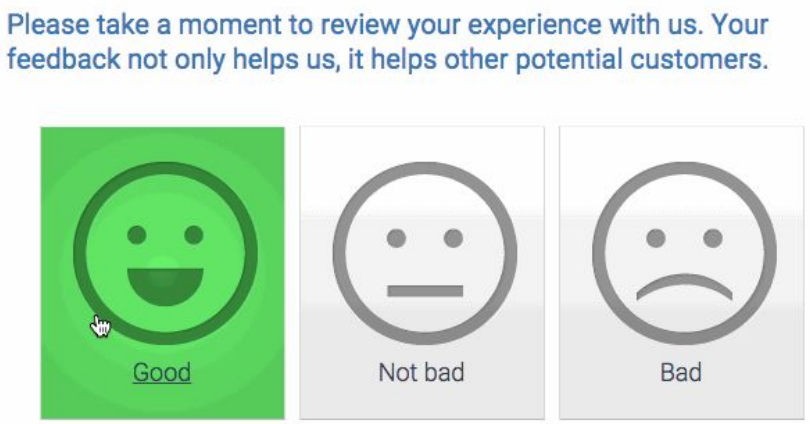


RESULTS



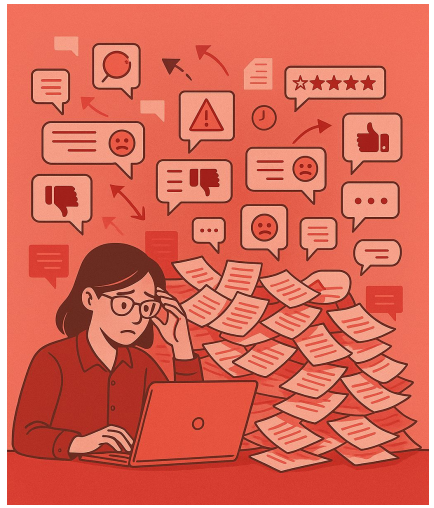
TAKEAWAYS

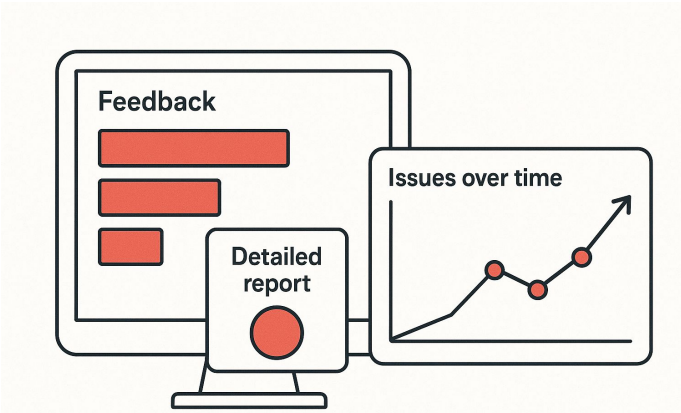
Background and Motivation



TITLE: Just Okay
REVIEW TEXT: I try to get Keratin treatments every 3 months, but honestly it has been getting costly. So, when I saw this I was excited to try it. I found it difficult to use and almost impossible to get to saturate the back of my hair and straight iron it the way they do in the salon. .....I will resume my regular treatments at my salon.",

Background and Motivation





Background and Motivation

Yes

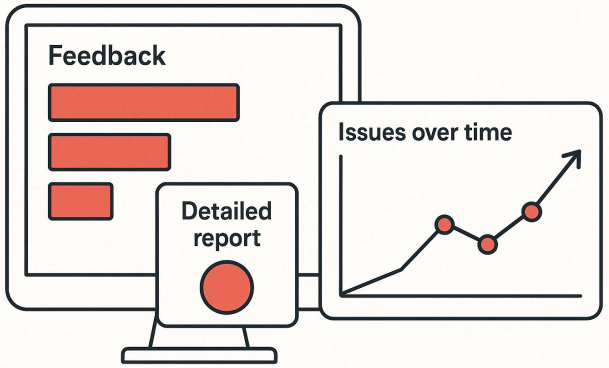
Okay

Zero

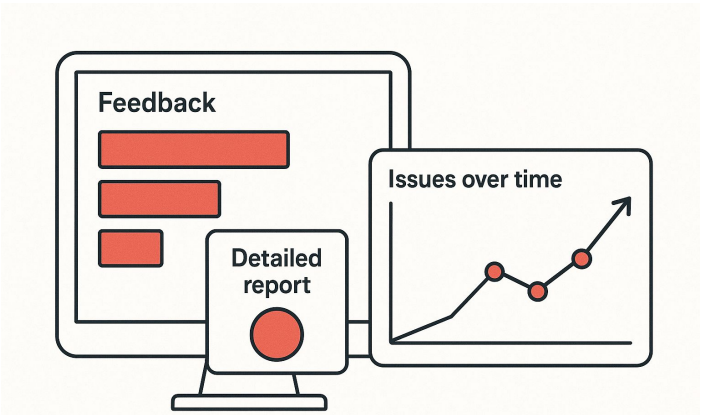
Good

Detailed description
....
....

Not two adequate



Background and Motivation



- Context window constraints
- Black box summaries
- Inaccurate insights
- Different results for different runs

Goal

- A transparent and scalable framework combining topic modeling with LLM prompting.
- An empirical analysis of different topic generation strategies.

Dataset

	Original Response	Synthesized Rewrite	Combined (Original + Synthesized)
Description	Direct transcription from customer audio, unmodified	LLM summarized and clarified rewrite of the original feedback	Raw feedback paired with its corresponding rewrite for richer context
Example	'good', 'okay'	'Customer found the troubleshooting page helpful.'	'okay' (raw), 'Customer found troubleshooting page helpful' (synthesized)

Approach

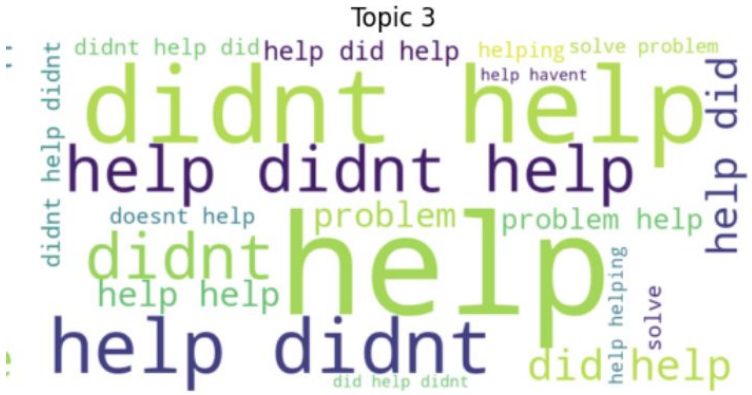
Topic Modelling

Topic Assignment

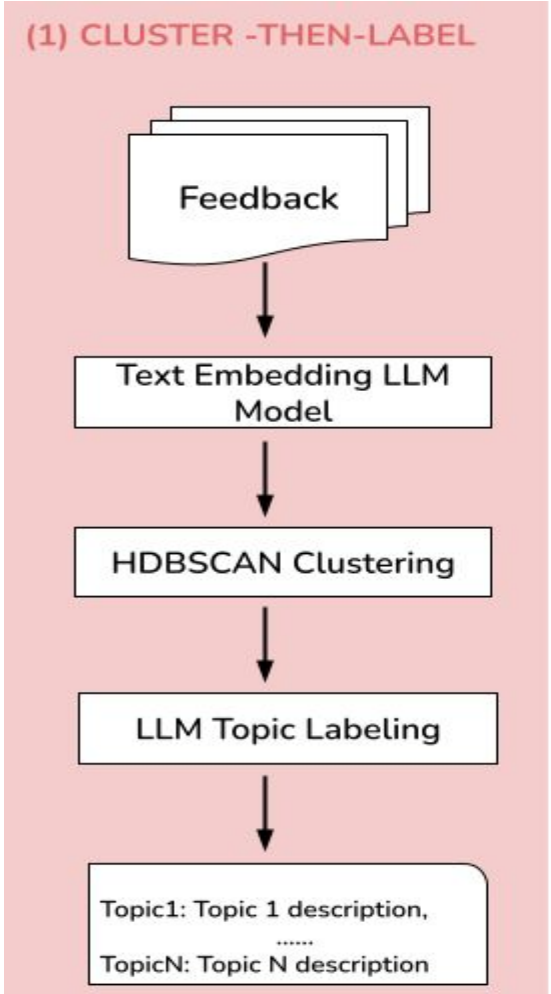
Insight Generation

Topic Modeling

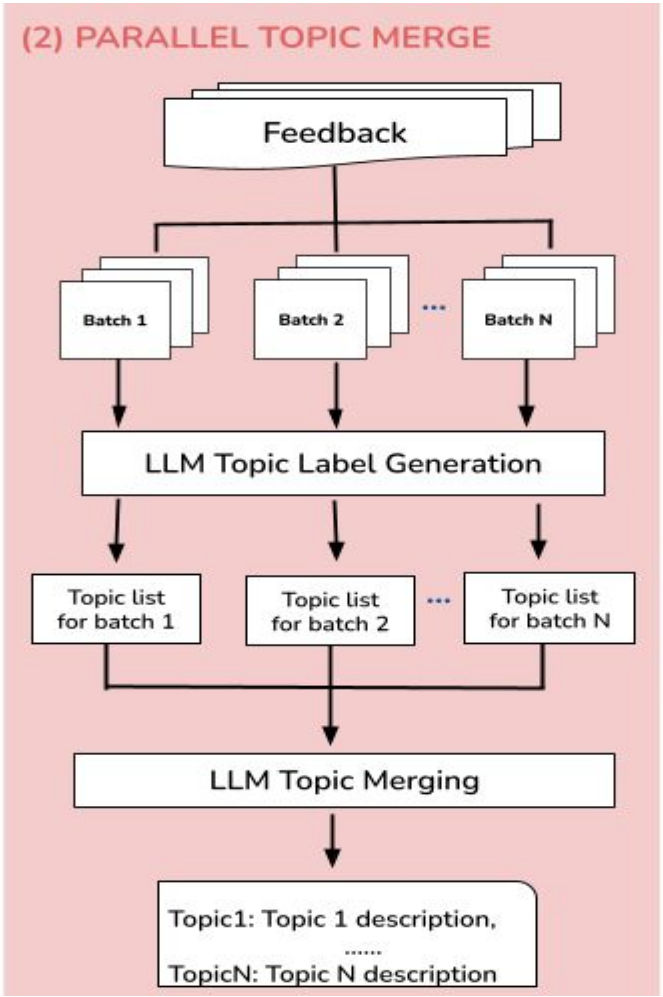
Discovers hidden themes by automatically **clustering** words that frequently appear together.



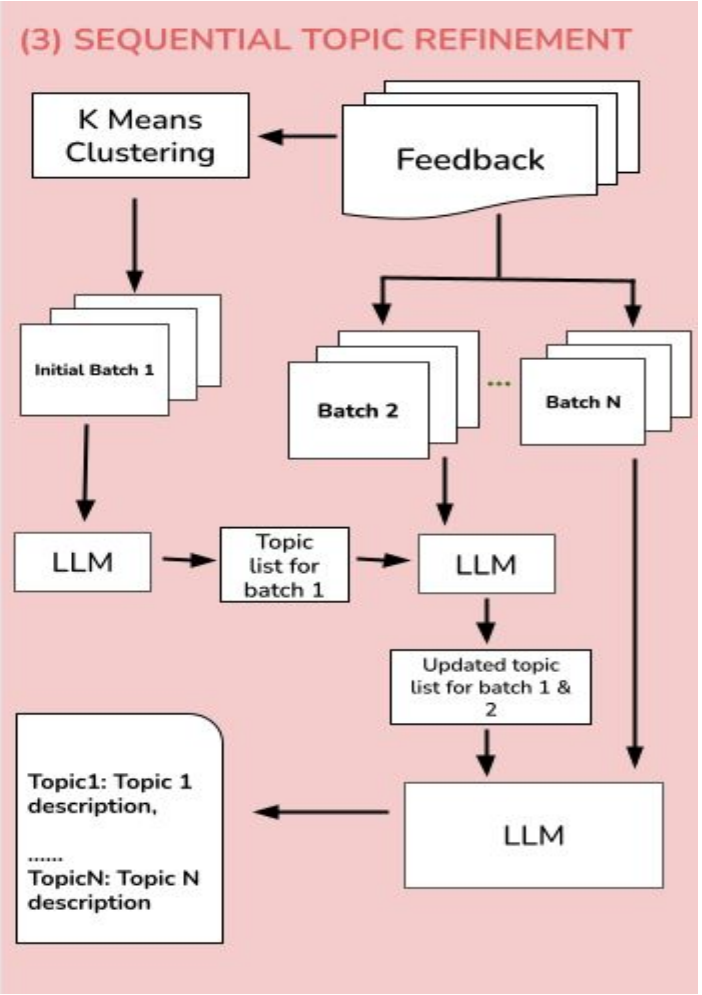
Topic Modeling



Topic Modeling



Topic Modeling

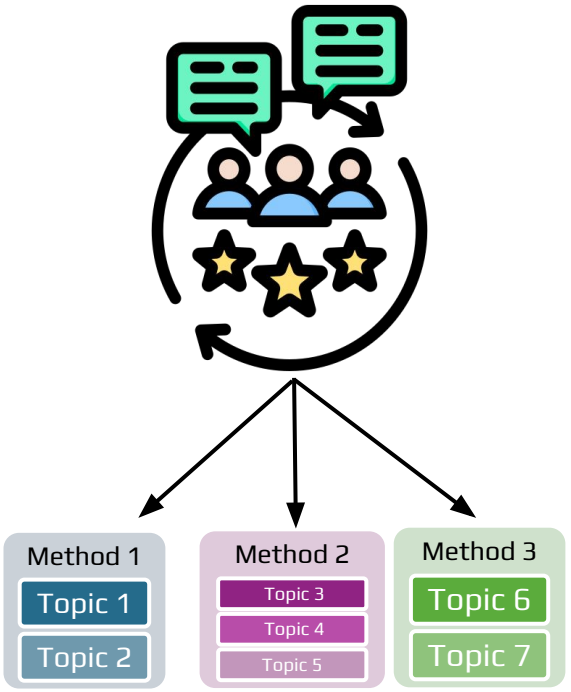


Phase 1: Topic Modeling

Phase 2: Topic Labeling

Phase 3: Label Aggregation

Phase 4: Scoring

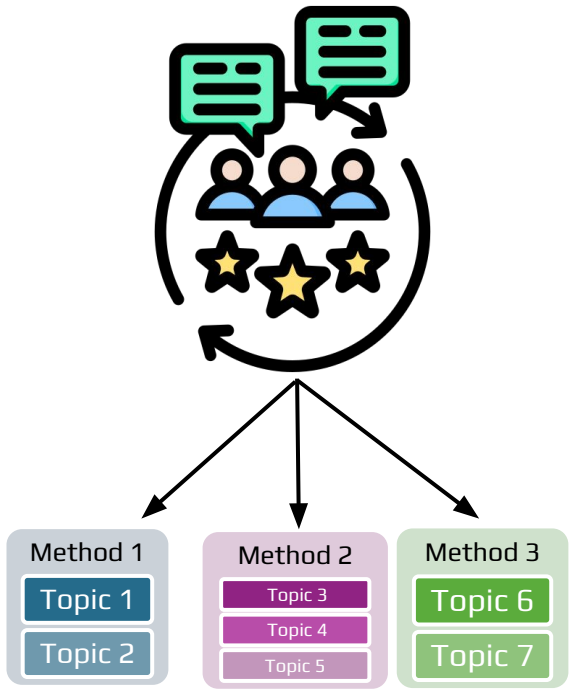


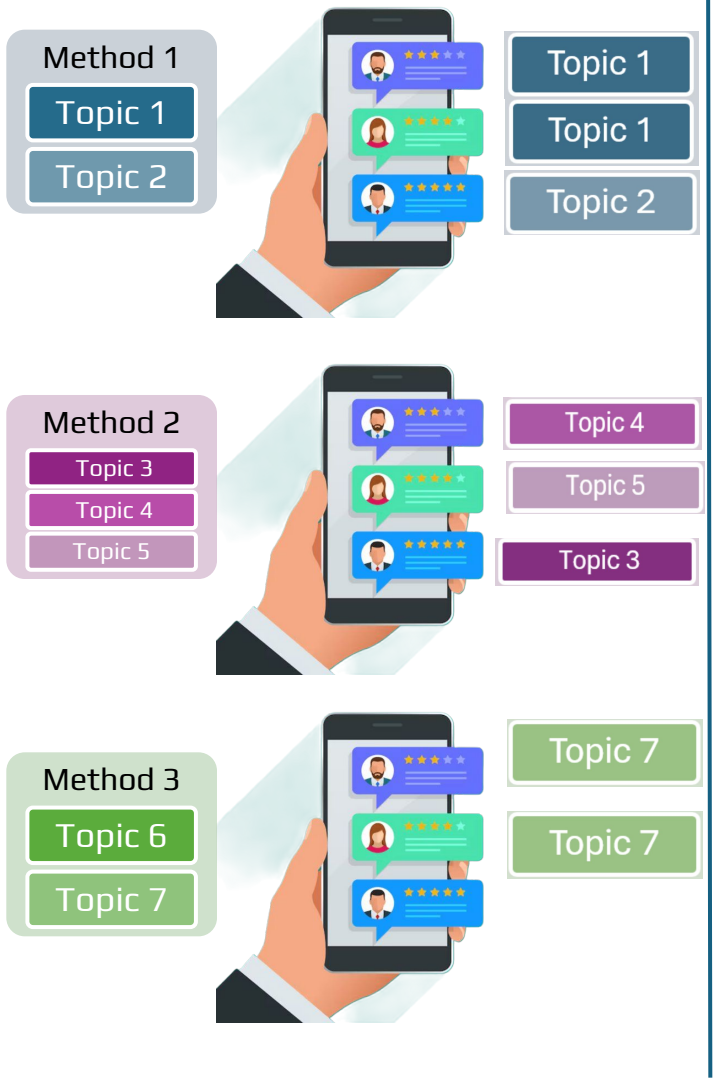
Phase 1: Topic Modeling

Phase 2: Topic Labeling

Phase 3: Label Aggregation

Phase 4: Scoring



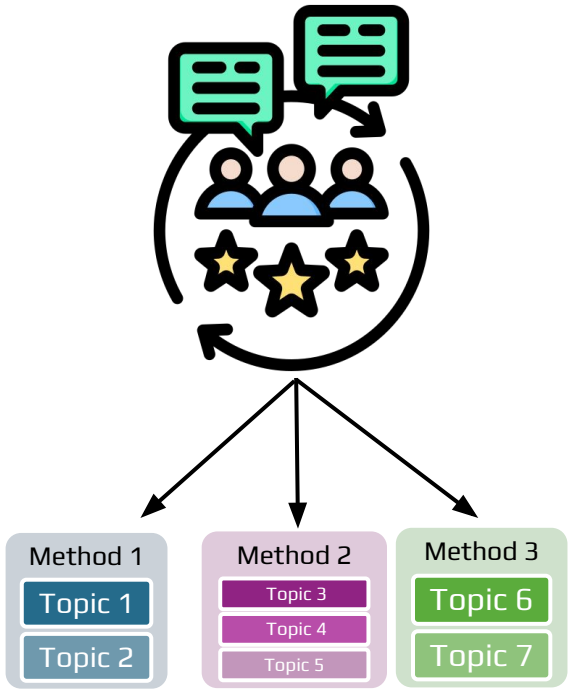


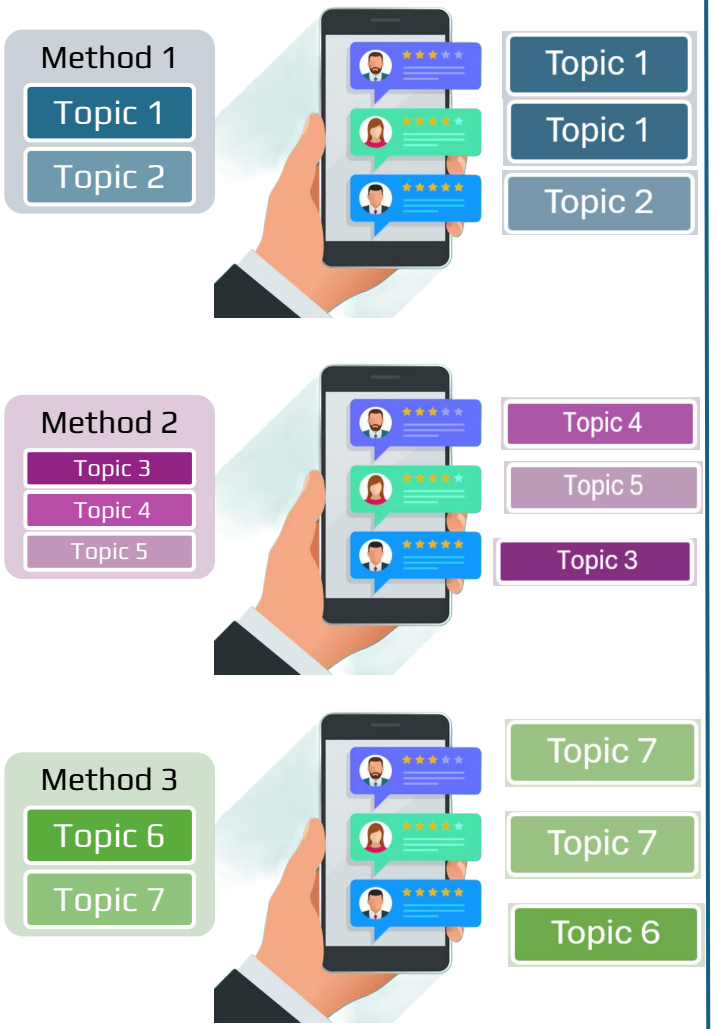
Phase 1: Topic Modeling

Phase 2: Topic Assignment

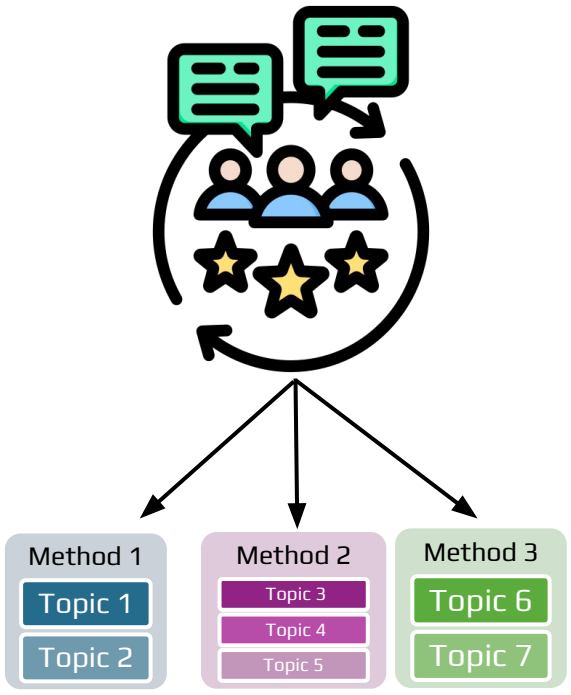
Phase 3: Label Aggregation

Phase 4: Scoring

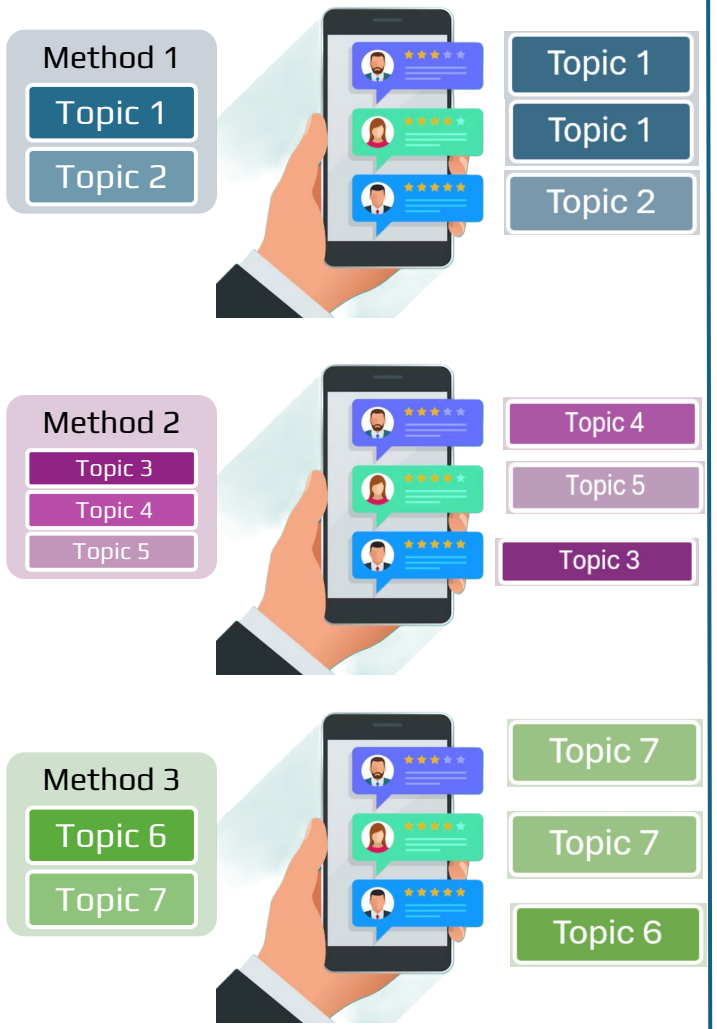




Phase 1: Topic Modeling



Phase 2: Topic Labeling



Phase 3: Label Aggregation

Phase 4: Scoring

Evaluation

FEEDBACK: I could not connect with it at all.

- Topic generated by Method 1: Could Not Access Troubleshooting
- Topic generated by Method 2: Automated Service Quality Evaluation
- Topic generated by Method 3 : Internet Connection Issues

Cosine Similarity

Textual Entailment

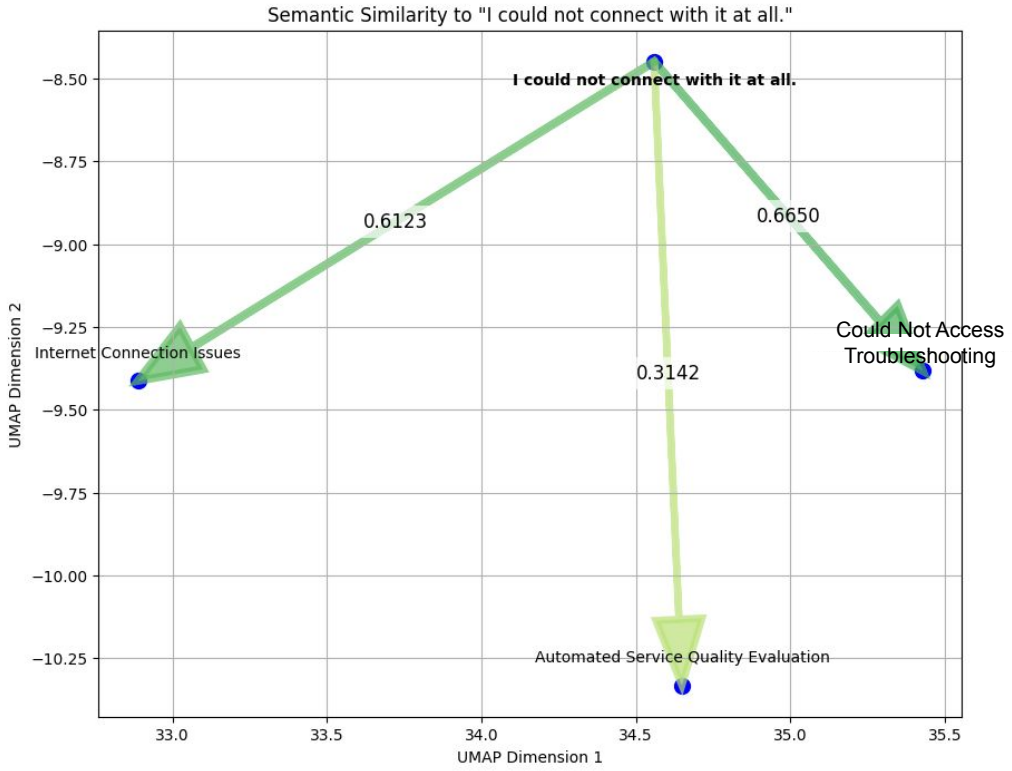
LLM Preference Evaluation

Manual Human Evaluation

Evaluation: Cosine Similarity

FEEDBACK: I could not connect with it at all.

- Topic generated by Method 1: Could Not Access Troubleshooting
- Topic generated by Method 2: Automated Service Quality Evaluation
- Topic generated by Method 3 : Internet Connection Issues



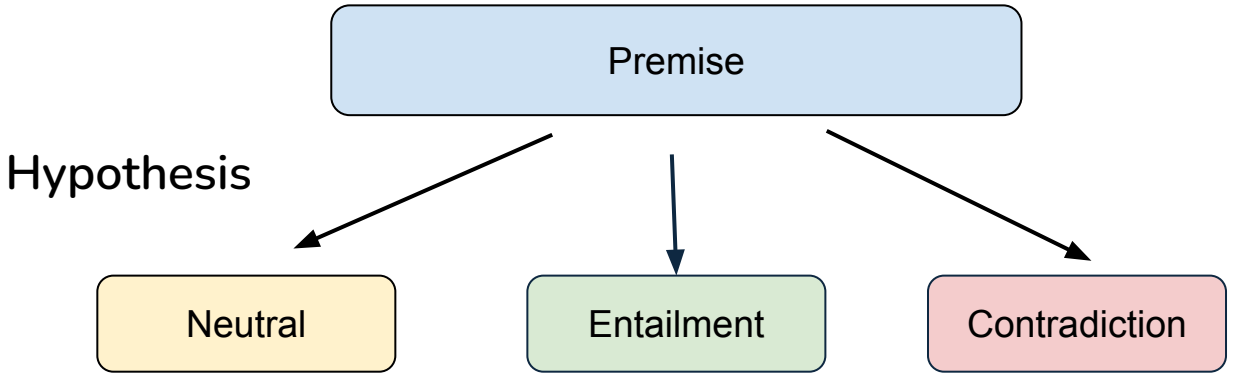
Evaluation: Textual Entailment

FEEDBACK: I could not connect with it at all.

- Topic generated by Method 1: Could Not Access Troubleshooting
- Topic generated by Method 2: Automated Service Quality Evaluation
- Topic generated by Method 3 : Internet Connection Issues

Entailment : One statement logically and necessarily follows from another. If the first statement is true, the second statement *must* also be true.

Premise: I saw a *grey hound* at the dog park.
Hypothesis: I saw a *dog* at the dog park.



Evaluation: Textual Entailment

FEEDBACK: I could not connect with it at all.

- Topic generated by Method 1: Could Not Access Troubleshooting
- Topic generated by Method 2: Automated Service Quality Evaluation
- Topic generated by Method 3 : Internet Connection Issues

Entailment : One statement logically and necessarily follows from another. If the first statement is true, the second statement *must* also be true.

Premise: "Could Not Access Troubleshooting"
Hypothesis: "I could not connect with it at all."

Entailment: 99.2% 🏆
Neutral: 0.8%
Contradiction: < 0.1%

Evaluation: LLM Preference Selection

FEEDBACK: I could not connect with it at all.

- Topic generated by Method 1: Could Not Access Troubleshooting
- Topic generated by Method 2: Automated Service Quality Evaluation
- Topic generated by Method 3 : Internet Connection Issues

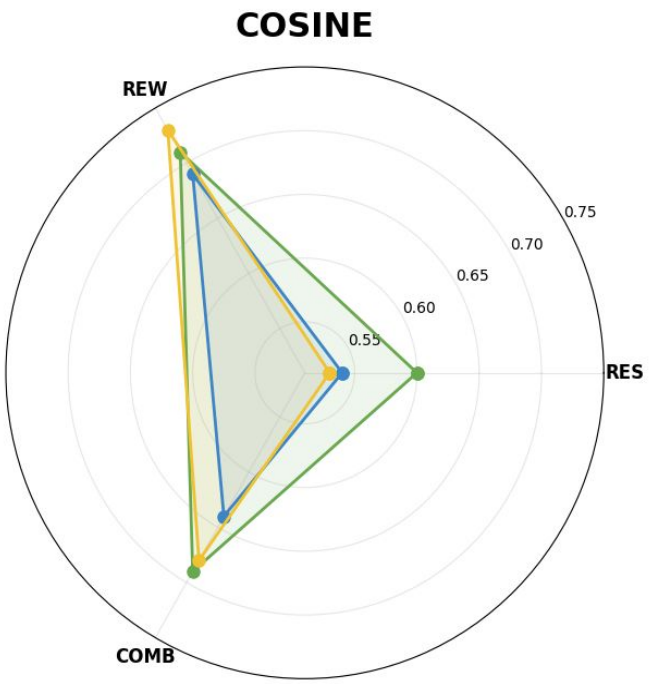
$$\%\text{Preference}_M = \frac{\sum_{f \in F} [C(f) = L_M(f)]}{|F|} \times 100\%$$

Where:
f: A single feedback being evaluated.
L_M(f): The Label from System M.
C(f): The Choice made by the judge.

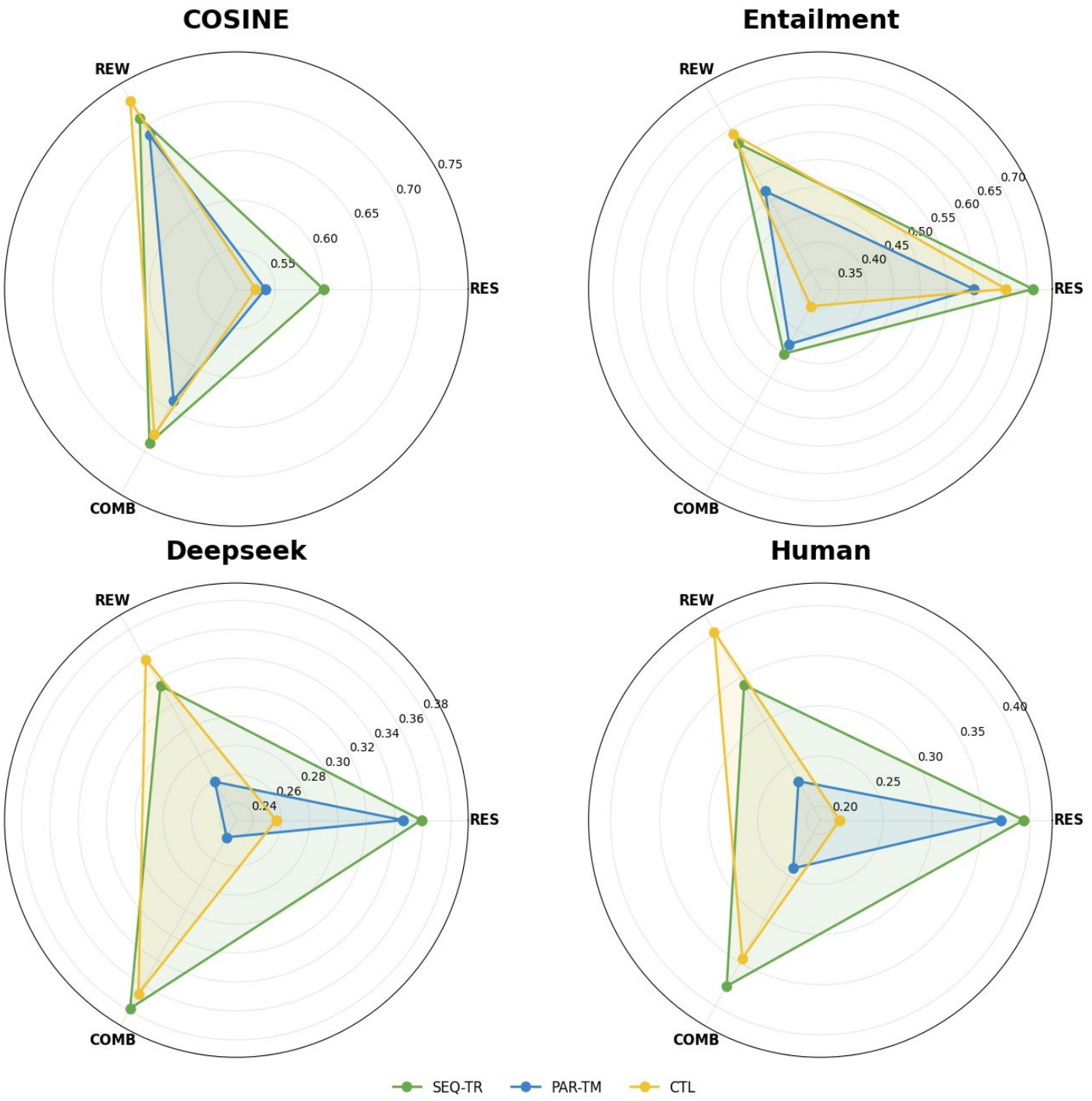
Experiment 1: Which method is better?

Method	Original Response (RES)	Synthesized Rewrite (REW)	Combined (COMB)
CTL			
SEQ-TR			
PAR-TM			

Results:
Which method is better?



Results:
Which method is better?



Results: What data input is best for a given method?

Evaluator	Method	RES	REW	COMB	N/A
Human	CTL	0.19	0.30	**0.39**	0.12
	PAR-TM	**0.42**	0.22	0.29	0.07
	SEQ-TR	**0.44**	0.25	0.28	0.03
DeepSeek	CTL	0.26	0.31	**0.34**	0.09
	PAR-TM	**0.46**	0.24	0.17	0.13
	SEQ-TR	**0.37**	0.28	0.27	0.08

Results: What data input is best for a given method?

Evaluator	Method	RES	REW	COMB	N/A
Human	CTL	0.19	0.30	**0.39**	0.12
	PAR-TM	**0.42**	0.22	0.29	0.07
	SEQ-TR	**0.44**	0.25	0.28	0.03
DeepSeek	CTL	0.26	0.31	**0.34**	0.09
	PAR-TM	**0.46**	0.24	0.17	0.13
	SEQ-TR	**0.37**	0.28	0.27	0.08

Example 1:

Original response (RES): Yeah

Synthesized Rewrite (REW): Customer found troubleshooting helpful

SEQ_{RES}: Miscellaneous
SEQ_{REW}): Helpful Troubleshooting
SEQ_{COMB}: Helpful Troubleshooting

PAR_{RES}: Out-of-Scope Feedbacks
PAR_{REW}: Positive Troubleshooting Experience
PAR_{COMB}: Helpful Troubleshooting Assistance

Example 2:

Original response (RES): No good.

Synthesized Rewrite (REW): Customer found troubleshooting link to be no help.

$HDBSCAN_{RES}$: Didn't Solve Problem
$HDBSCAN_{REW}$: Unhelpful Troubleshooting Page
$HDBSCAN_{COMB}$: Unhelpful Troubleshooting Feedback

Generating Insights

1. Quantifying Topic Importance and Sentiment
Measure each topic’s share of total feedback to gauge its prominence.
Assign sentiment labels (positive, neutral, negative) emotional tone of feedback.

2. Tracking Trends and Emerging Issues
Monitor growth trends by tracking topic share over time and fitting a trend line.
Identify whether topics are growing, declining, or stable.

3. Prioritizing Topics for Action
Combine topic share, sentiment profile, and trend direction into an impact scoring framework.

Takeaways

1. Using topic modeling and assignment helps generate transparent insights.
2. SEQ-TR excel on Original Response and Combined due to its sequential, context-aware prompting.
3. CTL, benefits from richer semantic embeddings in Synthesized Rewrites.

QUESTIONS?
THANK YOU