GenAI Insights from Insurance Call Transcripts

Sentiment & Call Outcome Detection using Cohere LLM

Presented by *Alice*
2028-05-13

Project Overview

Goal:

Analyze member transcripts to determine:

- Sentiment (Positive, Neutral, Negative)
- Call Outcome (Issue Resolved, Follow-Up Needed)

Scope:

- 200 call transcripts
- Focus on **customer ("Member:")** lines only

Why It Matters

Business Value:

- Track customer satisfaction at scale – Compared to Questionnaire (high cost, inefficient)
- Spot unresolved issues needing escalation – Compared to manually logged and follow-up (slow & high missing rate)
- Improve agent training with real feedback – Compared to standalone performance evaluation system (high cost)

Benefits:

- Data-driven insights
- Proactive issue resolution
- Better customer experience
- Reduce operational cost

Before and After using GenAI

Use Case	**Before GenAI**	**After GenAI**
Customer complaints	Detected manually (random QA)	Automatically flagged in real time
Agent coaching	Based on occasional reviews	Continuous feedback from every call
Monthly reporting	High effort, slow	Automated summaries, weekly insights
Escalation Tracking	Manual logs	NLP-driven resolution tagging

Data Source & Assumptions

Data:

- 200 `.txt` files, each a full call transcript
- Only "Member:" lines used for analysis

Assumptions:

- Transcripts are accurate and representative
- Speaker tags are consistent
- LLM outputs reflect customer sentiment

Tech Stack

Environment: (easy to handle)

- Python (in Google Colab) – Excellent for DS to experiment
- Cohere LLM API (`command-r`) – abundant of GenAI model for selection, without the need to hold the models, which will be expensive

Libraries: (generic toolkits)

- `pandas`, `cohere`, `matplotlib`, `seaborn`, `tqdm`

Pipeline Overview:

1. Load transcripts
2. Extract member lines
3. Use GenAI models for sentiment analysis & call-outcome classification
4. Aggregate and visualize

Key Results

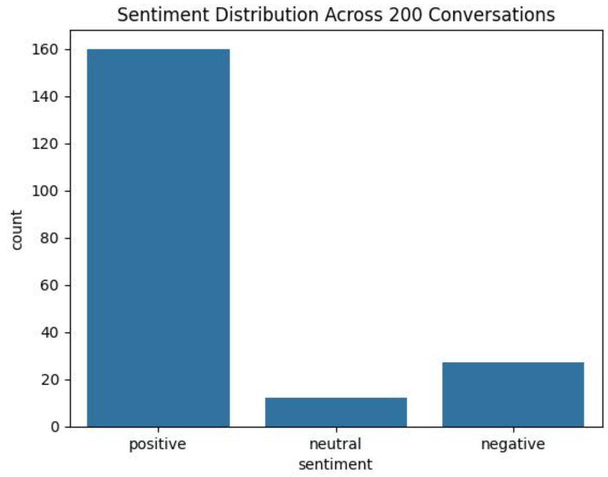
📊 **Sentiment Distribution**

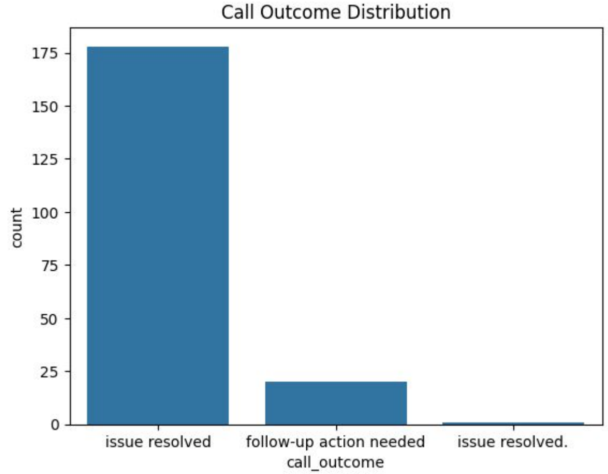
📊 **Call Outcome Distribution**

🧮 **Cross Tab:**

- **89.5%** issues **resolved** on-spot – Mostly with **positive** sentiment
- **10.5%** issues need **follow-up actions** – mostly with **negative** sentiment

call-outcomes	negative	neutral	positive	Grand Total
follow-up action needed	8.00%	1.50%	1.00%	10.50%
issue resolved	5.50%	5.00%	79.00%	89.50%
Grand Total	**13.50%**	**6.50%**	**80.00%**	**100.00%**





call-outcomes	negative	neutral	positive	Grand Total
follow-up action needed	76.19%	14.29%	9.52%	100.00%
issue resolved	6.15%	5.59%	88.27%	100.00%
Grand Total	**13.50%**	**6.50%**	**80.00%**	**100.00%**

- For cases need for a follow-up: **14.29% 'neutral' can be improved to 'Positive'**
- For issue solved cases: it make sense to **investigate the 6% 'negative' and 5.6% 'neutral**

Sentiment	follow-up action needed	issue resolved	Grand Total
positive	1.25%	98.75%	100.00%
neutral	23.08%	76.92%	100.00%
negative	59.26%	40.74%	100.00%
Grand Total	**10.50%**	**89.50%**	**100.00%**

- For 'neutral' clients: **investigate why 76% with issue-resolved but still not happy ('positive')**
- For 'negative' clients: investigate why 40% with **issue-resolved but still angry ('negative')**

Risks & Limitations

⚠️ **Model Risks:**

- Hallucinations / misclassification
- Ambiguity in customer tone

📉 **Data Risks:**

- Inconsistent formatting
- Missing speaker labels
- Potential transcription errors

🧪 **Mitigation:**

- Manual review sampling
- Flag uncertain outputs

Future Improvements

🔁 **Next Steps:**

- Fine-tune Cohere model for higher accuracy
- Include agent lines for fuller context
- Automate real-time monitoring of calls
- Include mega data for call transcript to understand more information, such as handling time

🧠 **Advanced Ideas:**

- Emotion tracking
- Conversation summarization
- Escalation prediction

Thank You