

With our flagship *DataInFormationSM* suite of solutions for Image Labeling, Data Annotation, NLP Calibration and Form Data Capture, Liberty Source is proud to serve a wide range of clients across many industry verticals. Our *Solution Briefs* provide a snapshot of the challenges clients bring to us, and the outcomes we achieve for them.

Audio Transcription & Validation

Industry	Administrative	Data Type	Audio
Project Duration	30 Days	Ongoing?	Yes

Challenge

DataInFormation was tasked with transcribing and providing context for audio recordings from various corporate meetings. This cannot be done using automation transcription exclusively because of the complexity of the audio: information would be left out or incorrectly transcribed when topics are complex and nuanced, and when participants speak simultaneously.

Solution

Data Associates listened to audio files of varying length and validated the initial machine transcription of the conversation to ensure the capture of all information, including overlapping audio segments, as well as annotating specific data points needing to be captured.

Outcome

Accurate audio transcription was collected and made available for reference, review, and compliance purposes. The transcribed output was then used as ground-truth training for audio transcription machine learning models.

Audio Transcription & Validation

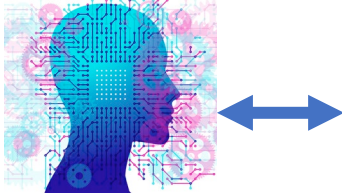
Raw Audio Data



'First Pass' Annotation of Audio Transcription

Segment	Speaker	Text
1	Shankar Vedantam	This is Hidden Brain.
2	Em Shankar Vedantam	
4		One spring afternoon in 2011, Jeremy Clifton was in a rush.
5		He was living in Atlanta, trying to make ends meet, working as a tutor, a guitar teacher and fitness instructor.
6		On that day, he was running late to one of his many gigs.
7		But as he waited for a train on the subway platform ...
8	Jer. Clifton	

Human in the Loop annotates and verifies accuracy of audio transcription



JSON Output

```
annotated_audio - Notepad
File Edit Format View Help
{"version": "1.1", "data": {"audio": [{"content": "Shankar Vedantam", "tokens": ["Shankar", "Vedantam"], "metadata": {}}, {"content": "This is Hidden Brain platform ...", "tokens": ["But", "as", "he", "waited", "for", "a", "train", "on", "the", "subway", "platform", "..."], "metadata": {}}, {"content": "Jer C Shankar Vedantam", "tokens": ["Shankar", "Vedantam"], "metadata": {}}, {"content": "The third rail, of course, is the live rail.", "tokens": ["at sparks flew up and down his body, and I thought he was dead.", "tokens": ["I", "was", "just", "six", "feet", "above", "it", "and", "trying", "to", "pull", "it", "up", "out", "of", "the", "subway", "tracks.", "metadata": {}}, {"content": "Shankar Vedantam", "tokens": ["Shankar", "Vedantam"], "metadata": {}}, {"content": "Jeremy Clifton joins us right now, live in the beautiful city of Atlanta, Georgia", "tokens": ["Shankar Vedantam", "tokens": ["Shankar", "Vedantam"], "metadata": {}}, {"content": "He didn't like the way the story was unfolding. In to ask a question that was fundamentally psychological in nature: Why was his perception of the incident so different?", "tokens": ["what", "clip", "tokens": ["News", "clip", "metadata": {}}, {"content": "... with the bombing of PLO bases and installations in and around the Lebanese to two groups.", "tokens": ["I", "do", "not", "know", "later", "the", "psychologist", "Lee", "Ross", "showed", "media", "coverage", "of", "the", "incident", "as", "ata", "were", "really", "quite", "remarkable.", "tokens": ["The", "data", "were", "really", "quite", "remarkable.", "metadata": {}}, {"content": "There were no id, no, the clips were depicting Palestinians in a bad light.", "tokens": ["The", "Palestinians", "said", "no", "the", "clips", "were", "depicting", "coverage", "to", "be", "seen", "as", "fair", "and", "impartial.", "tokens": ["The", "materials", "were", "designed", "by", "CBS", "and", "NBC", "News", "who", "if", "anything", "e", "two", "groups", "had", "seen", "two", "different", "realities.", "metadata": {}}, {"content": "Lee Ross eventually came up with a term for this bias, and others, who", "tokens": ["and", "others", "who", "are", "deeply", "invested", "in", "a", "particular", "ideology.", "metadata": {}}, {"content": "Lee Ross", "tokens": ["Shankar", "Vedantam"], "metadata": {}}, {"content": "Lee Ross, who died in 2021, conducted a number of other experiments to show how peopleivity", "tokens": ["that", "I", "have", "a", "sense", "of", "what", "I", "perceive", "is", "the", "way", "it", "really", "is.", "metadata": {}}, {"content": "And t", "tokens": ["Vedantam"], "metadata": {}}, {"content": "This is why when others don't agree with us on some important issue, we readily assume there must", "tokens": ["metadata": {}}, {"content": "One says, 'The room is too cold.'", "tokens": ["One", "says", "'The", "room", "is", "too", "cold.'"], "metadata": {}}, {"content": "George", "tokens": ["Carlin", "had", "a", "line.", "metadata": {}}, {"content": "He said, 'Did you ever notice that?'", "tokens": ["he", "said", "'Did", "you", "ever", "notice", "that?'"], "metadata": {}}, {"content": "He was studying to become a psychologist.", "tokens": ["he", "was", "studying", "to", "become", "a", "psychologist.", "metad", "e", "felt", "his", "behavior", "on", "the", "subway", "platform", "was", "not", "unusual", "because", "he", "tended", "to", "have", "an", "optimistic", "view", "of", "the", "game", "he", "was", "playing.", "because", "he", "wanted", "to", "write.", "metadata": {}}, {"content": "Shankar Vedantam", "tokens": ["Sh", "with", "emotional", "disorders.", "tokens": ["he", "was", "studying", "people", "with", "emotional", "disorders.", "metadata": {}}, {"content": "Jer Clift", "tokens": ["that", "these", "extremely", "negative", "beliefs", "were", "a", "part", "of", "depression", "somehow", "but", "they", "thought", "that", "they", "were", "under", "if", "our", "beliefs", "might", "actually", "shape", "the", "world", "we", "see.", "tokens": ["Ross", "of", "us", "think", "that", "we", "too", "the", "world", "is", "let", "re", "presented as a passive-aggressive comment instead of a simple reminder of an upcoming deadline.", "tokens": ["An", "ambiguous", "email", "from", "a"]}]}
```