

Subjective-QA

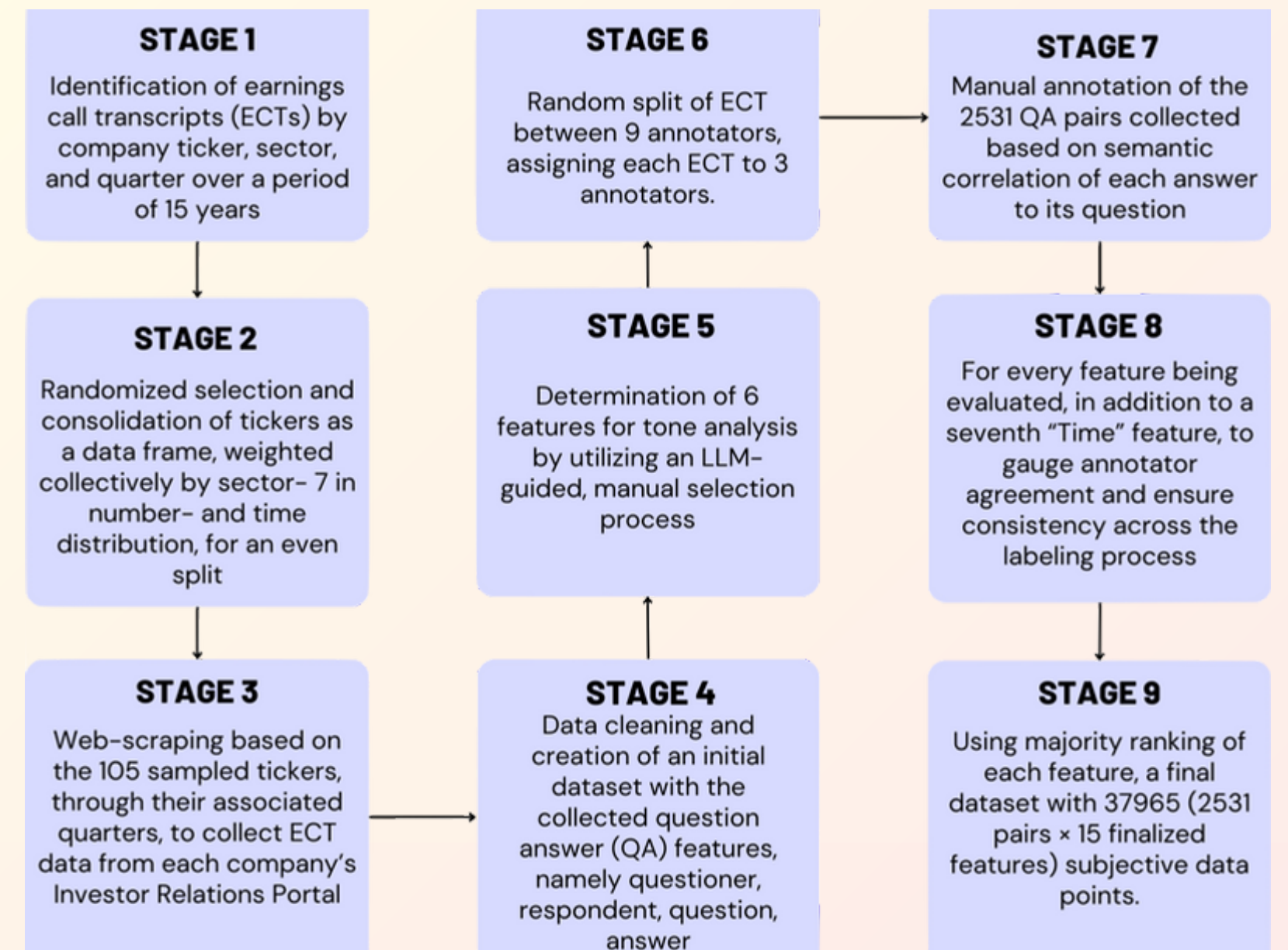
Huzaifa Pardawala , Veer Kejriwal , Siddhant Sukhani, Abhishek Pillai, Rohan Bhasin, Dhruv Adha, Tarun Mandapati, Andrew DiBiasio, Agam Shah, Sudheer Chava

Financial Services & Innovation Lab, Georgia Institute of Technology, Atlanta, GA 30332

Summary

A number of datasets for sentiment classification and stock price forecasting have been created using Earnings Call Transcripts, additionally aiding downstream tasks such as sentiment analysis, numerical claim detection, and hawkish-dovish-neutral sentiment classification. However, none of these datasets focus on the multiple dimensions of answers given during the QA sessions of Earnings Calls. To address this limitation, we propose SubjECTive- QA, the first annotated dataset of answers given during earnings calls. We will also provide an empirical evaluation with various state-of-the-art Pre-trained Language Models as benchmarks for the tone classification task in the future and highlight the limitations posed by our dataset. Our dataset is strategically designed to forecast headwinds or tailwinds for organizations. Our codebase is currently private.

Methodology



Annotation Guidelines

We employed Microsoft Excel for the annotation procedure. The annotators were asked to strictly adhere to the following annotation guide: Give the answer a rating of:

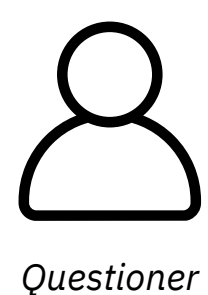
- **2:** If the answer positively demonstrates the chosen feature, with regards to the question.
- **1:** If there is no evident/neutral correlation between the question and the answer for the feature.
- **0:** If the answer negatively correlates to the question on the chosen feature.

At the end, the individual annotations were combined based on the majority rating. In case there was no clear majority that particular rating was assigned the value '1'.

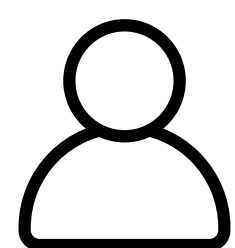
Feature	Description
Relevant	In context of the question, the speaker has answered the question appropriately.
Specific	The speaker includes appropriate technical details in the answer.
Cautious	The speaker discusses taking a more conservative, risk-averse approach.
Assertive	The speaker is certain of what is happening/what is going to happen.
Clear	The speaker is transparent about what they want to convey.
Optimistic	The speaker hopes for a positive outcome in the future.

Table 1: Description of the chosen features

Sample Annotation

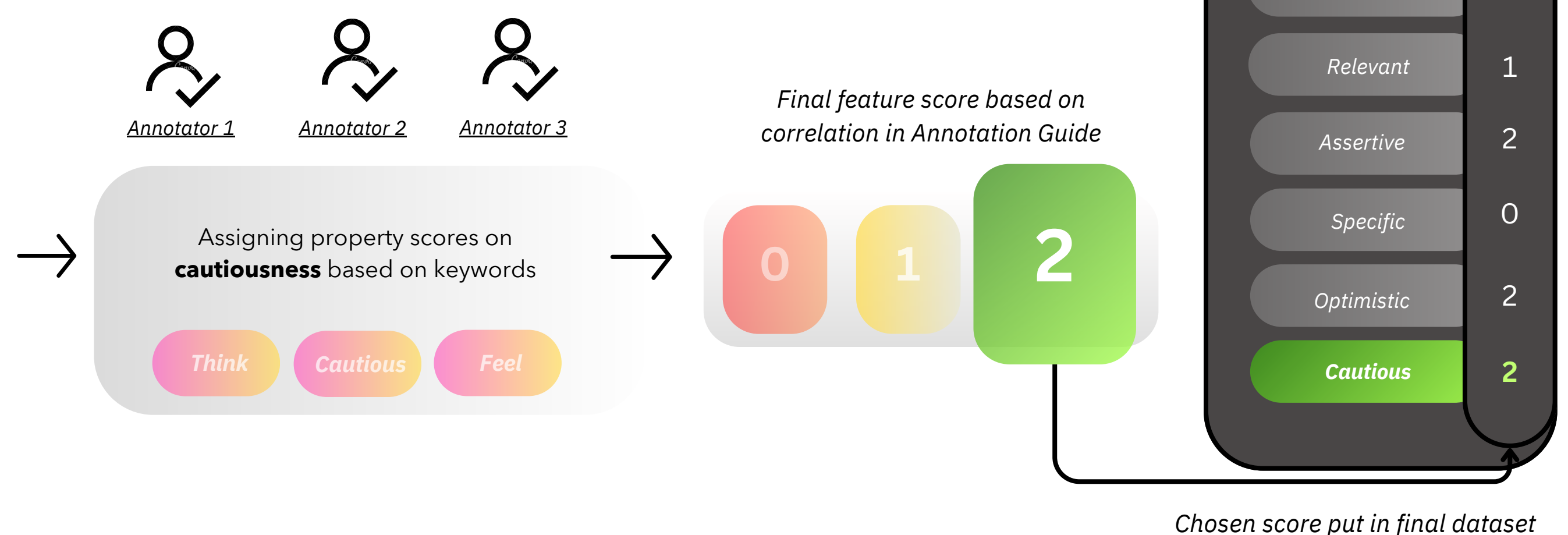


And then, just wanted to ask also, with the sharp drop in prices over the last couple of months ... sharp decline over the last two months?



To some degree, I almost feel like we're victims of our own actions. I think as prices start to fall, everybody gets a little more cautious on buying ... is going to be

Manual Annotation Process



Annotator Agreement

To address the low Krippendorff's alpha scores, we added a 7th feature, Time, which we score as follows:

- **2:** if the answer discusses only future events
- **1:** if the answer pertains to the present time, or if multiple time frames are mentioned
- **0:** if the answer discusses only past events

The objective "Time" feature is anticipated to have a Krippendorff's alpha score, indicating that prior annotator disagreements were likely due to subjective features, not inaccuracies.

Feature	Aggregate Krippendorff's alpha
Relevant	0.197
Specific	0.288
Cautious	0.233
Assertive	0.214
Clear	0.206
Optimistic	0.205

Table 2: Krippendorff's Alpha

Results

- Respondents were much more likely to respond optimistically than being cautious. Respondents were **2.9x** more likely to answer a question optimistically.
- For answers with a score of 0, showing no satisfaction with a certain probability, specificity was the highest with **20%** of all answers not being specific, indicating the variance in the quality of answers.
- Another interesting result we derived was of the neutral tonal nature of the answers, with **70%** and **59%** of the answers receiving a cautious and optimistic score of 1 respectively, indicating that respondents answered most questions neutrally.