Automated pop-up fact-checking works by converting audio of a live event into text, locating the claims therein, and applying matching fact-checks done by reputable organizations—all in real time. Our team focused on improving the matching accuracy for this automated process. Given a claim to match, our system works by first identifying a list of candidate fact-checks in a database, and then performing a more expensive NLP-based test on each pair of claim and candidate fact-check. At the beginning of our project, although the system brought up a multitude of matches, many weren’t ones we wanted to show to the users, due to errors stemming from sentence fragmentation in audio transcription, inadequate handling of numerical quantities, and general lack of relevancy. We set out to increase matching accuracy by curating data and establishing measures for evaluation, and by improving various steps in our processing pipeline—from audio transcription, identification of candidate fact-checks, to NLP-based testing.

1. Use human-harnessed data for training
2. Improve audio transcription of claims
3. Increase accuracy of results from SOLR
4. Strengthen comparison score with BERT

The final stage of the automated pipeline compares each pair of a claim and a candidate fact-check identified by Solr, using a natural language processing model. Previously, the pipeline used a Transformer model to perform this step but we explored using the current state-of-the-art model called BERT which improves upon the transformer model. Neither model has seen much high-quality training data related to our goals and purposes, thus reducing the quality of our results.

The BERT model takes in the claim and fact checked statement and outputs a score quantifying how similar the sentences are. On its own, BERT outperforms the previous transformer model however it still does not perform as well as we hope. We hope that by using our own data to finetune the model will increase performance. We are also exploring ways to combine the two models to provide more accurate results by only showing matches whose BERT and Transformer scores meet set thresholds. For the precision and recall values reported the threshold for the independent transformer and BERT scores were .85 and when combined .85 for BERT and .17 for the transformer.