

Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster

*Zeyu Zhang, Jacob Daniel Devasier, Damian Jimenez, Sai
Sandeep Naraparaju, Haiqi Zhang, Zhengyuan Zhu, Chengkai Li*

Falsehoods in Elections

- **Many forces tried interfere with the elections**
 - One video featured a Haitian man (although he was not really Haitian) saying he had just gotten to the United States and had voted in two counties—Gwinnett and Fulton—in Georgia, but it turned out to be a fake video made in Russia. <https://www.bbc.com/news/articles/c9vnyl2jnpjo>
- **100+ active fact-checking sites (PolitiFact.com, FullFact.org, CNN, Washington Post, ...)**

Combating Falsehoods in Elections

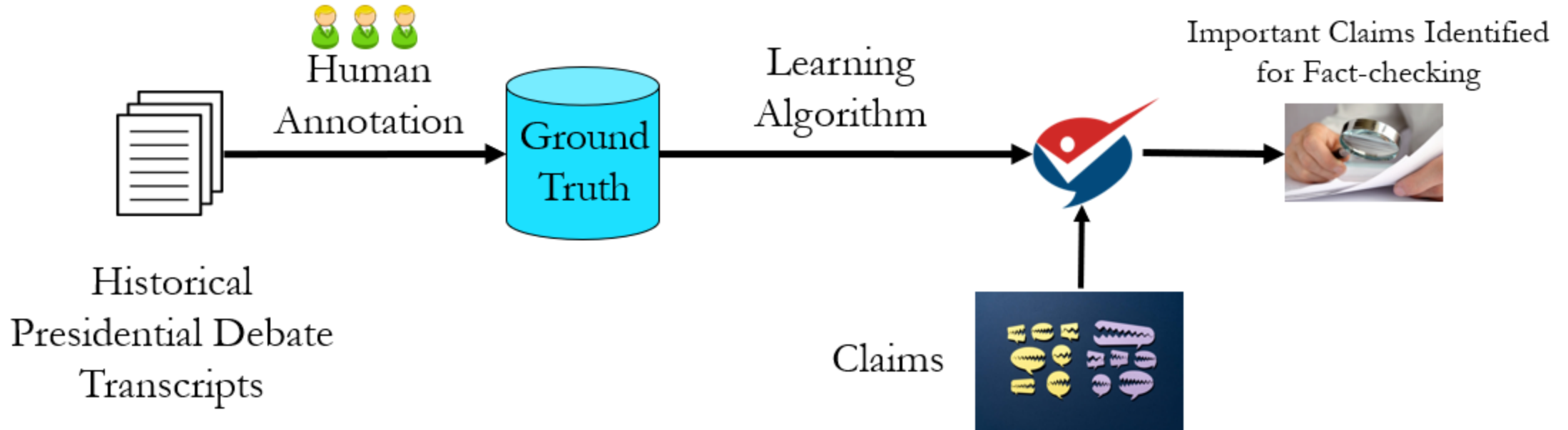
- **Massive claims made by news articles, social media posts, speeches of public figures, etc.**
- **How to identify the factual claims worth fact-checking?**
 - Check-worthy: Video shows people in Los Angeles looting from homes during wildfires!!!
 - Non-Check-worthy: I love Los Angeles. It is so beautiful!

AI Models to Detect Check-worthy Claims

- **Train machine learning models to identify check-worthy factual claims**
- **ClaimBuster**
 - <https://idir.uta.edu/claimbuster>
 - One of the earliest and most practical models
 - Contributed to at least 33 claims featured in 30 articles by prominent fact-checking outlets, including The Washington Post
 - The API currently has 362 registered users ranging from academics to professional journalists across several countries.
 - The API has been called over 219 million times
 - Fact-checkers and others can use ClaimBuster during the live coverage of an event to help them discover claims that need more caution to trust.

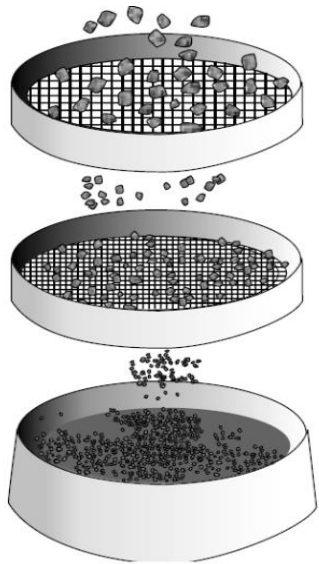
How Does ClaimBuster Work?

- **Given a statement, ClaimBuster outputs a score ranging from 0 to 1. Scores closer to 1 indicate high check-worthiness, while scores closer to 0 indicate low check-worthiness.**
- **ClaimBuster is trained on historical presidential debates**



Ground Truth Dataset


- We implemented a sophisticated annotation mechanism to ensure the quality
- Data Annotation website: https://idir.uta.edu/classifyfact_survey
- Dataset: <https://zenodo.org/records/3836810>
- Codebase: <https://github.com/utaresearch/claimbuster-spotter>



25000+ sentences
400+ annotators
90000+ labels

100 top-quality annotators
63000+ labels

Majority voting
23000+ admitted sentences


ClaimBuster
naeemulhassan labeled 562 sentences
Leaderboard
Instructions
Log Out

Your estimated total payment is \$37.39. You are paid approximately 6.70¢ per sentence on average. Your total payment is ranked 8 out of 54 participants. Click 'Leaderboard' to see details. This message and the leaderboard are updated every 15-30 sentences. [See tips for improving your pay rate at the bottom of this page.](#)

RO: Now, I voted for the Patriot Act.

[More Context](#)

Will the general public be interested in knowing whether (part of) this sentence is true or false?

☐ There is **no** factual claim in this sentence.
 ☐ There is a factual claim but it is **unimportant**
☒ There is an **important** factual claim.

ClaimBuster on Presidential Debates

- We apply ClaimBuster on each year's presidential debates, the transcripts are posted on the ClaimBuster website, with each sentence annotated with its check-worthiness score

2024 Final Presidential Debate

September 10, 2024

Philadelphia, PennsylvaniaNational Constitution CenterABC

Transcript Source: <https://abcnews.go.com/Politics/harris-trump-presidential-debate-transcript/story?id=113560542>

Moderators: David Muir, Linsey Davis

Participants: Donald Trump, Kamala Harris

Chronological Order

or

Order by Score

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

David Muir: President Trump, I'll give you two minutes.

Donald Trump: First of all, I have no sales tax. That's an incorrect statement. She knows that. We're doing tariffs on other countries. Other countries are going to finally, after 75 years, pay us back for all that we've done for the world. And the tariff will be substantial in some cases. I took in billions and billions of dollars, as you know, from China. In fact, they never took the tariff off because it was so much money, they can't. It would totally destroy everything that they've set out to do. They've taken in billions of dollars from China and other places.

Case Study: 2024 Presidential Debates

- **We examined two CNN articles covering some fact-checks of the presidential debates held on June 27 and September 10, 2024**
 - <https://www.cnn.com/2024/06/27/politics/fact-checking-the-cnn-presidential-debate/index.html>
 - <https://www.cnn.com/2024/09/10/politics/fact-check-debate-trump-harris/index.html>
- **Of the 71 claims selected and fact-checked by CNN from these debates, the average ClaimBuster score was 0.69, significantly higher than the average score of 0.41 for all sentences in the debates.**

Case Study: 2024 Presidential Debates

- **Table 1 and 2 shows the number of sentences in the debate transcripts within specific ClaimBuster score ranges and how many of these sentences were fact-checked in the corresponding CNN articles.**

Table 1: Statistics of the Presidential Debate on June 27, 2024

Score Range	Sentence Count	Sentences Covered by CNN Article
0.8 - 1	59	9
0.6 - 0.8	216	18
0.4 - 0.6	438	9
0.2 - 0.4	539	1
0 - 0.2	167	0

Table 2: Statistics of the Presidential Debate on September 10, 2024

Score Range	Sentence Count	Sentences Covered by CNN Article
0.8 - 1	50	11
0.6 - 0.8	187	12
0.4 - 0.6	372	10
0.2 - 0.4	651	1
0 - 0.2	215	0

Thank You!