DEVELOPING AUTOMATED TOOLS FOR POLITICAL FACT-CHECKING

Duke UNIVERSITY | BASS CONNECTIONS

MATTHEW O' BOYLE[1], SHERRY FENG[1]

[1]Duke Department of Computer Science

Bass Connections in Information, Society & Culture

INTRODUCTION

The current process of fact-checking is, in a sense, "chaotic". In the current process, independent fact-checkers employ upwards of 20 fact-checkers to assess the authenticity of major claims made in a speech. Journalists then **filter** claims to assess what is ostensibly "checkworthy", and manually **search** through an archive of fact-checks to identify suitable matches. If a match exists, **commentary** about the claim gets written and a rating is published; if not, a reporter starts researching the claim, and then an editor revises and publishes it with a claim. This entire process can take a day at least, upwards of a week. Our team is automating the matching process in order to assist reporters who are otherwise manually seeking through the archives. This has the potential to improve efficiency by hours or even days, and moves us towards a more fair and transparent world for media.

DESIGN CHALLENGES

What do we optimize for when designing this web interface? The biggest challenge here was thinking about how we could design an interface that would be fast and intuitive. The two main challenges:

- **Speed:** since journalists are fact-checking in real time, we prioritized speed over all else, adding keyboard shortcuts and only offering the headlines of fact-checks to compensate, acknowledging that this came at the tradeoff of information clarity.
- **Concurrency:** since multiple fact-checkers could be performing the same fact-check at the same time, we need some way of indicating conflicting edits. UI features to solve this are listed to the right.

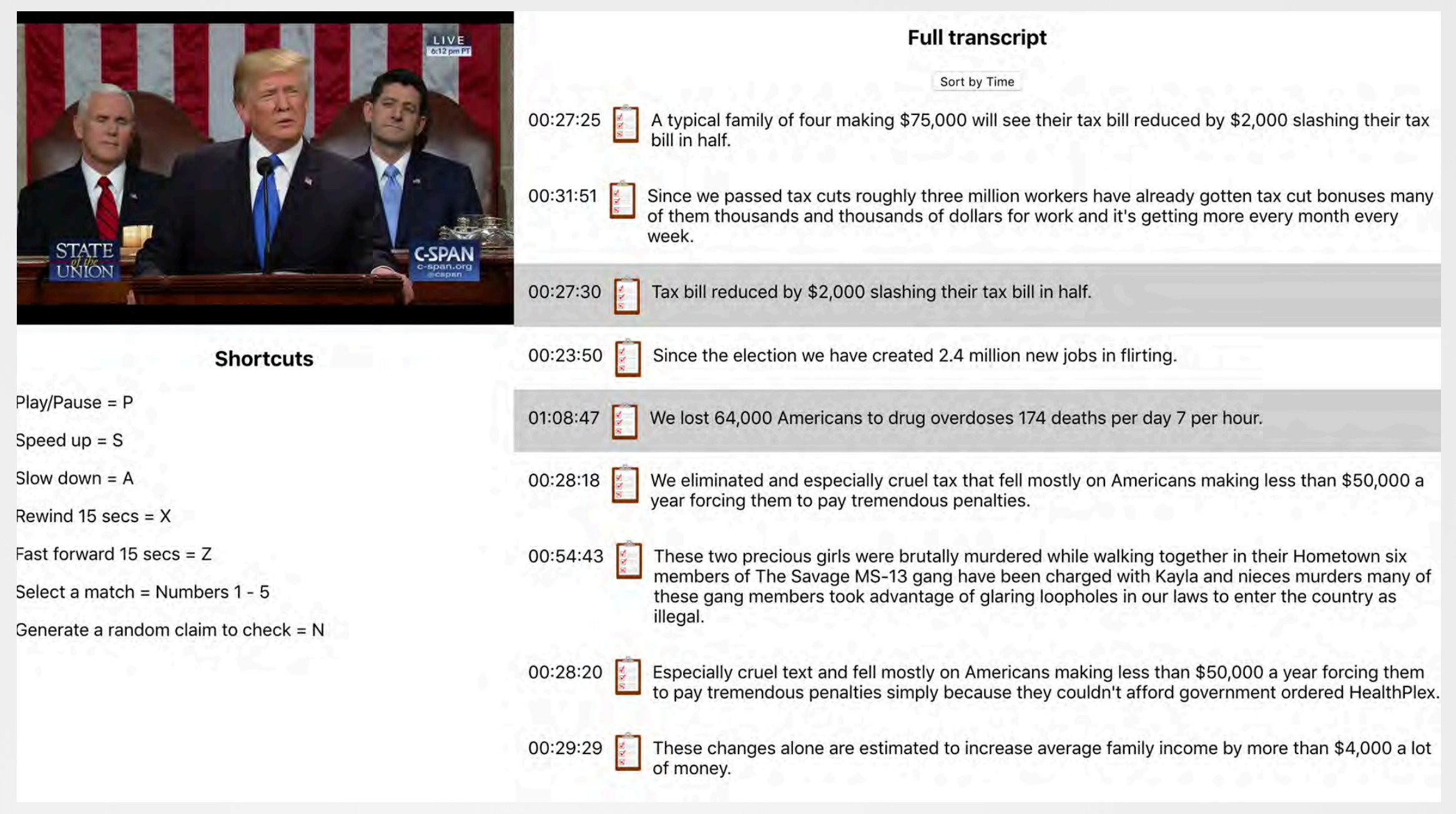
OUR INTERFACE FOR DATA VERTIFICATION

Our machine learning model, as is in its incipient stages, requires some human input to ensure the model is trained and accurate. We created a web interface in React and Ruby that allows journalists to match claims to fact-checks, and select fact-checks they want to present to the end user.

Clicking the time allows users to go back to that point in the video and verify the transcription

Keyboard shortcuts enable journalists to prioritize speed, crucial in live fact-checking

In three key presses, you could select a claim, pick its match, and return to the home screen to start the process again



Full transcript

Sort by Time

00:27:25	A typical family of four making $75,000 will see their tax bill reduced by $2,000 slashing their tax bill in half.
00:31:51	Since we passed tax cuts roughly three million workers have already gotten tax cut bonuses many of them thousands and thousands of dollars for work and it's getting more every month every week.
00:27:30	Tax bill reduced by $2,000 slashing their tax bill in half.
00:23:50	Since the election we have created 2.4 million new jobs in flirting.
01:08:47	We lost 64,000 Americans to drug overdoses 174 deaths per day 7 per hour.
00:28:18	We eliminated and especially cruel tax that fell mostly on Americans making less than $50,000 a year forcing them to pay tremendous penalties.
00:54:43	These two precious girls were brutally murdered while walking together in their Hometown six members of The Savage MS-13 gang have been charged with Kayla and nieces murders many of these gang members took advantage of glaring loopholes in our laws to enter the country as illegal.
00:28:20	Especially cruel text and fell mostly on Americans making less than $50,000 a year forcing them to pay tremendous penalties simply because they couldn't afford government ordered HealthPlex.
00:29:29	These changes alone are estimated to increase average family income by more than $4,000 a lot of money.

Shortcuts

Play/Pause = P

Speed up = S

Slow down = A

Rewind 15 secs = X

Fast forward 15 secs = Z

Select a match = Numbers 1 - 5

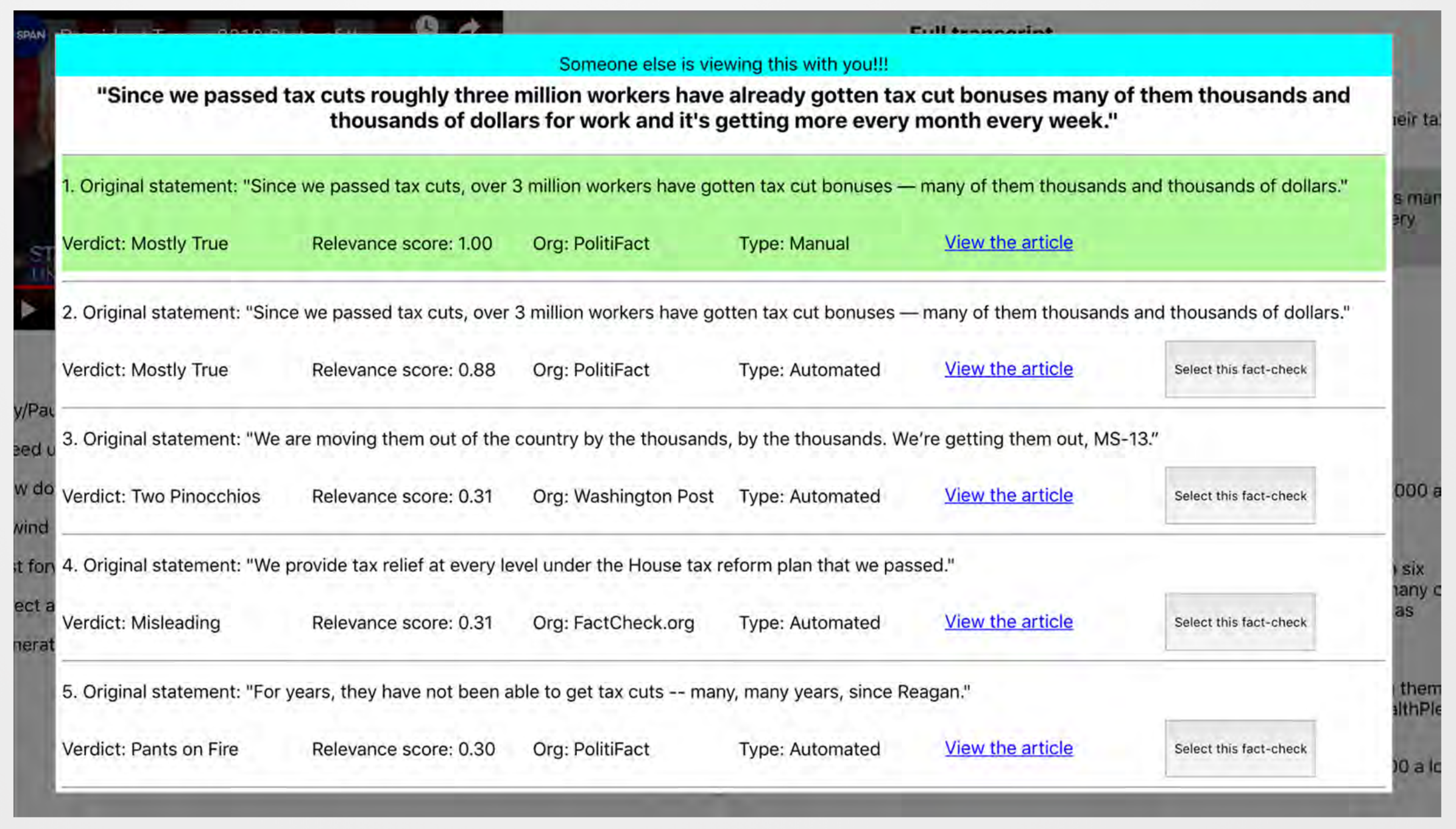
Generate a random claim to check = N

Buttons allow users to sort the transcript by time or checkworthiness, allowing journalists to prioritize their goals

Gray boxes highlight when other journalists are editing the stream; prevents concurrency conflicts

Green statements have already been manually verified by other journalists to be relevant to the claim

Keyboard shortcuts enable journalists to quickly select relevant fact-checks



Someone else is viewing this with you!!!

"Since we passed tax cuts roughly three million workers have already gotten tax cut bonuses many of them thousands and thousands of dollars for work and it's getting more every month every week."

1. Original statement: "Since we passed tax cuts, over 3 million workers have gotten tax cut bonuses — many of them thousands and thousands of dollars."
Verdict: Mostly True | Relevance score: 1.00 | Org: PolitiFact | Type: Manual | View the article

2. Original statement: "Since we passed tax cuts, over 3 million workers have gotten tax cut bonuses — many of them thousands and thousands of dollars."
Verdict: Mostly True | Relevance score: 0.88 | Org: PolitiFact | Type: Automated | View the article | Select this fact-check

3. Original statement: "We are moving them out of the country by the thousands, by the thousands. We're getting them out, MS-13."
Verdict: Two Pinocchios | Relevance score: 0.31 | Org: Washington Post | Type: Automated | View the article | Select this fact-check

4. Original statement: "We provide tax relief at every level under the House tax reform plan that we passed."
Verdict: Misleading | Relevance score: 0.31 | Org: FactCheck.org | Type: Automated | View the article | Select this fact-check

5. Original statement: "For years, they have not been able to get tax cuts -- many, many years, since Reagan."
Verdict: Pants on Fire | Relevance score: 0.30 | Org: PolitiFact | Type: Automated | View the article | Select this fact-check

Blue alert box highlights when other journalists are editing, preventing conflicts

Matches page allows journalists to view the original fact-check

Selecting the fact-check associates the claim as relevant to that listed fact-check

CHALLENGES

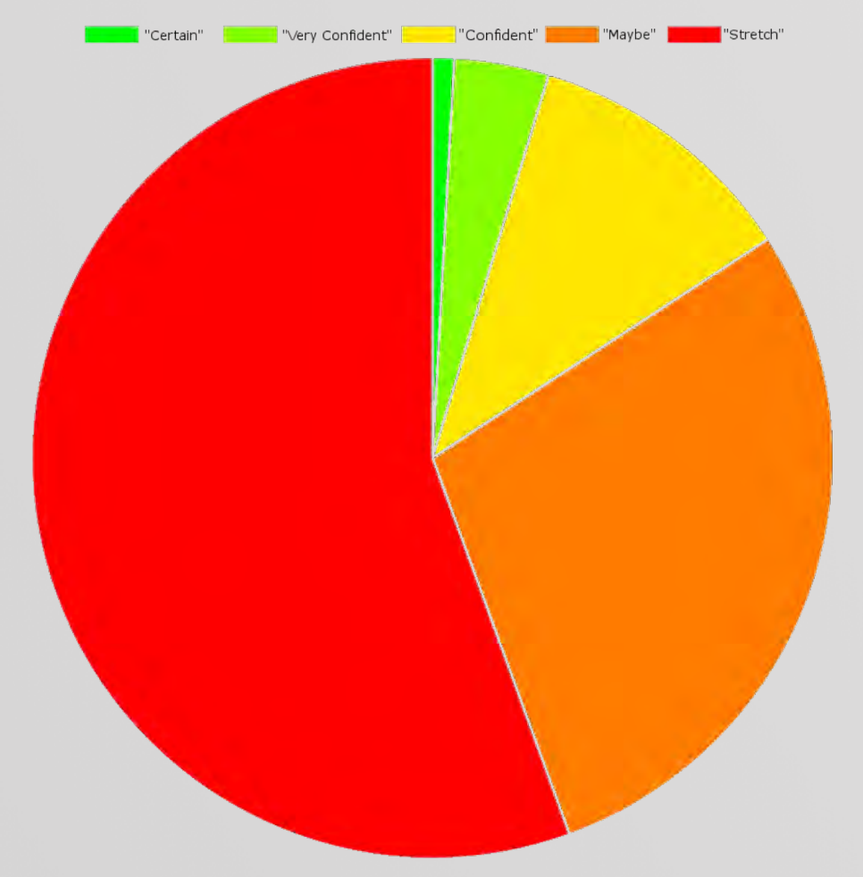
- Speed was the biggest challenge. To enable this, we only provide the headline from the fact-check article in our interface.
- However, this may not be enough information to make judgement calls, and providing a link to the article may result in lesser speed efficiency in the long term, since users would have to click to read.
- A secondary challenge was accuracy. When a user detects a mis-translation or mis-grouping, the updated transcription would have to run back through the pipeline, wasting valuable time in the process.

FUTURE GOALS

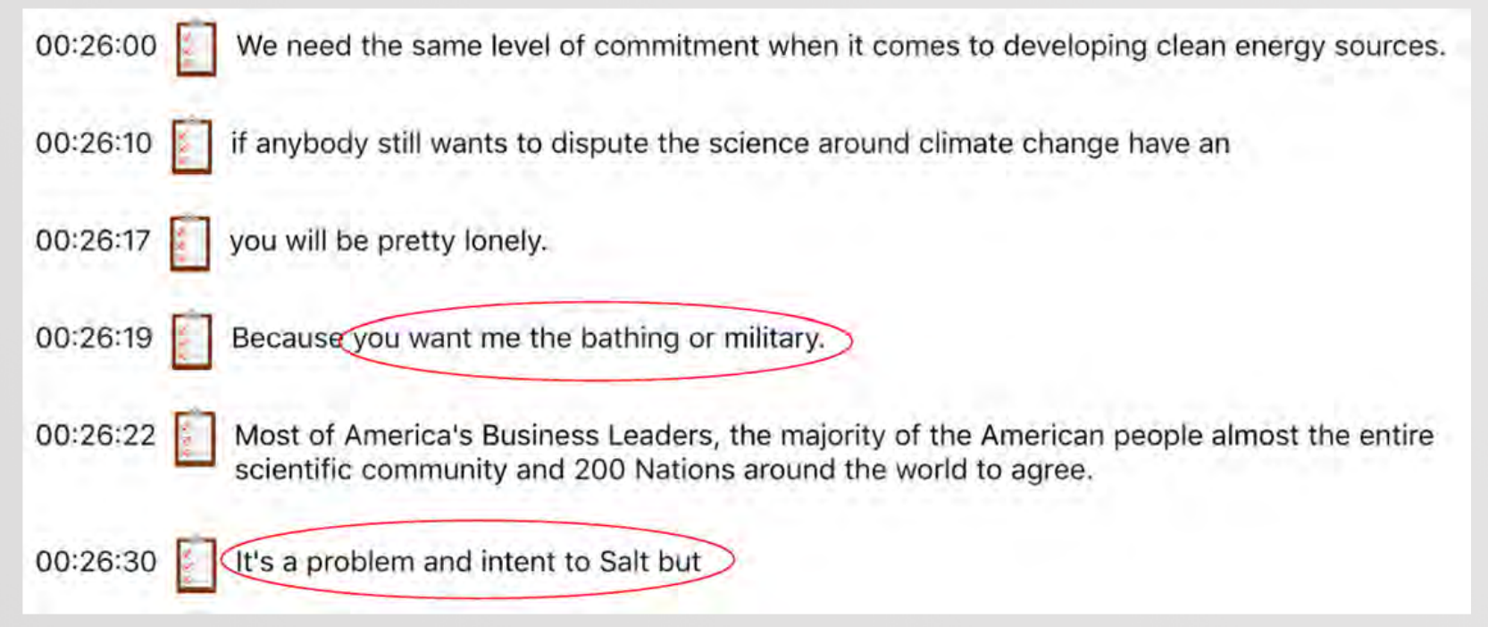
- **Search Feature.** To avoid transcription errors, we aim to allow users to input their own text to search for fact-checks if they feel the algorithm has mismatched or mistranslated.
- **Curator Optionality.** To provide more context to fact-checks and claims, we are developing options for curators to leave comments for other users and edit fact-checks before they are sent out to the end user, so that we can improve the gaps in information that arise from this speed tradeoff.

We hope to test this with real users during a live political event in the near future, which will allow us to fine-tune our user interface and create an even faster, more intuitive user interface.

PROBLEMS WITH AUTOMATION

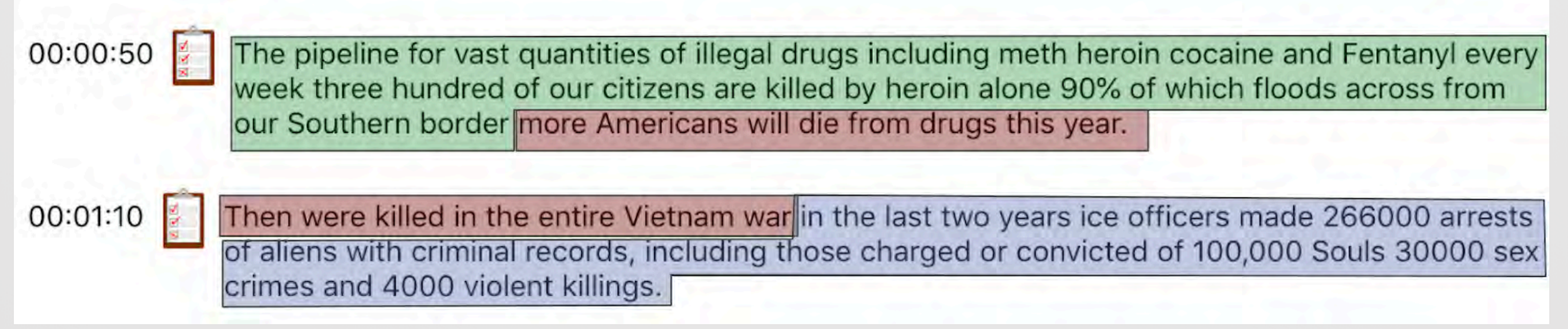


This chart shows the confidence of matches passed through our pipeline at 0.4 relevance



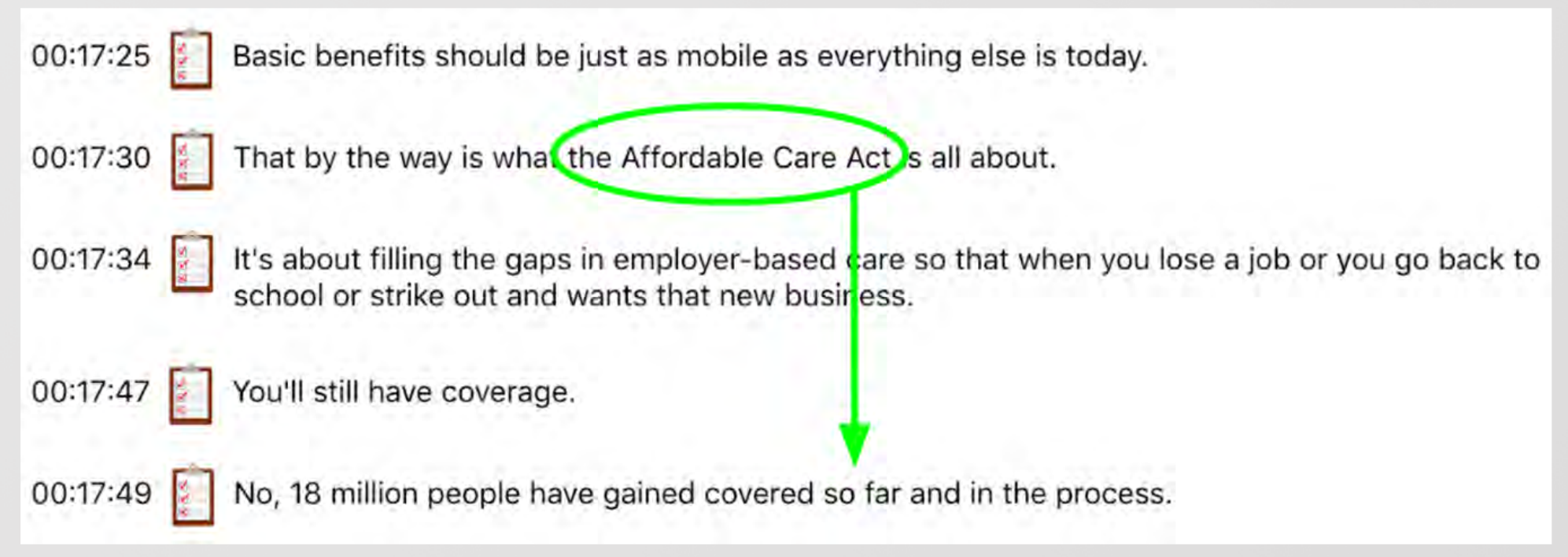
00:26:00	We need the same level of commitment when it comes to developing clean energy sources.
00:26:10	if anybody still wants to dispute the science around climate change have an
00:26:17	you will be pretty lonely.
00:26:19	Because you want me the bathing or military.
00:26:22	Most of America's Business Leaders, the majority of the American people almost the entire scientific community and 200 Nations around the world to agree.
00:26:30	It's a problem and intent to Salt but

Mistranslation: Google's speech-to-text "mishears" a sentence. This is heightened by applause and politicians talking over each-other



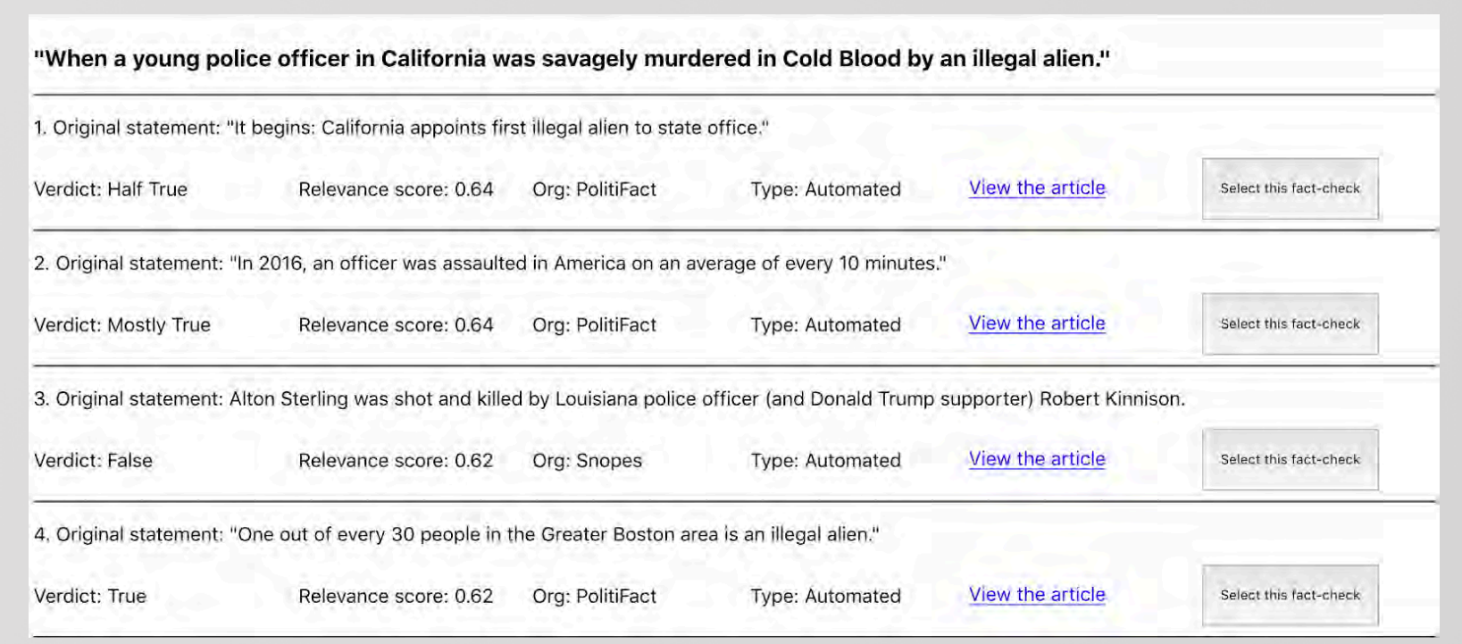
00:00:50	The pipeline for vast quantities of illegal drugs including meth heroin cocaine and Fentanyl every week three hundred of our citizens are killed by heroin alone 90% of which floods across from our Southern border more Americans will die from drugs this year.
00:01:10	Then were killed in the entire Vietnam war in the last two years ice officers made 266000 arrests of aliens with criminal records, including those charged or convicted of 100,000 Souls 30000 sex crimes and 4000 violent killings.

Incorrect Grouping: Our speech-to-text component tries to identify where sentences begin and end, but often splits check worthy sentences into two seemingly irrelevant parts



00:17:25	Basic benefits should be just as mobile as everything else is today.
00:17:30	That by the way is what the Affordable Care Act is all about.
00:17:34	It's about filling the gaps in employer-based care so that when you lose a job or you go back to school or strike out and wants that new business.
00:17:47	You'll still have coverage.
00:17:49	No, 18 million people have gained covered so far and in the process.

Difficulties with context: Our algorithm often fails to determine what "it"/"that" is referring to when a politician has mentioned a topic earlier in their speech



"When a young police officer in California was savagely murdered in Cold Blood by an illegal alien."

1. Original statement: "It begins: California appoints first illegal alien to state office."
Verdict: Half True | Relevance score: 0.64 | Org: PolitiFact | Type: Automated | View the article | Select this fact-check

2. Original statement: "In 2016, an officer was assaulted in America on an average of every 10 minutes."
Verdict: Mostly True | Relevance score: 0.64 | Org: PolitiFact | Type: Automated | View the article | Select this fact-check

3. Original statement: Alton Sterling was shot and killed by Louisiana police officer (and Donald Trump supporter) Robert Kinnison.
Verdict: False | Relevance score: 0.62 | Org: Snopes | Type: Automated | View the article | Select this fact-check

4. Original statement: "One out of every 30 people in the Greater Boston area is an illegal alien."
Verdict: True | Relevance score: 0.62 | Org: PolitiFact | Type: Automated | View the article | Select this fact-check

Mis-matching: Semantic similarities lead the algorithm to think a claim and fact-check are associated, despite them being completely different

Acknowledgments: This project was done under the Bass Connections project "Data and Technology for Fact-Checking", led by Jun Yang and Bill Adair.