

Data & Technology for Fact-Checking

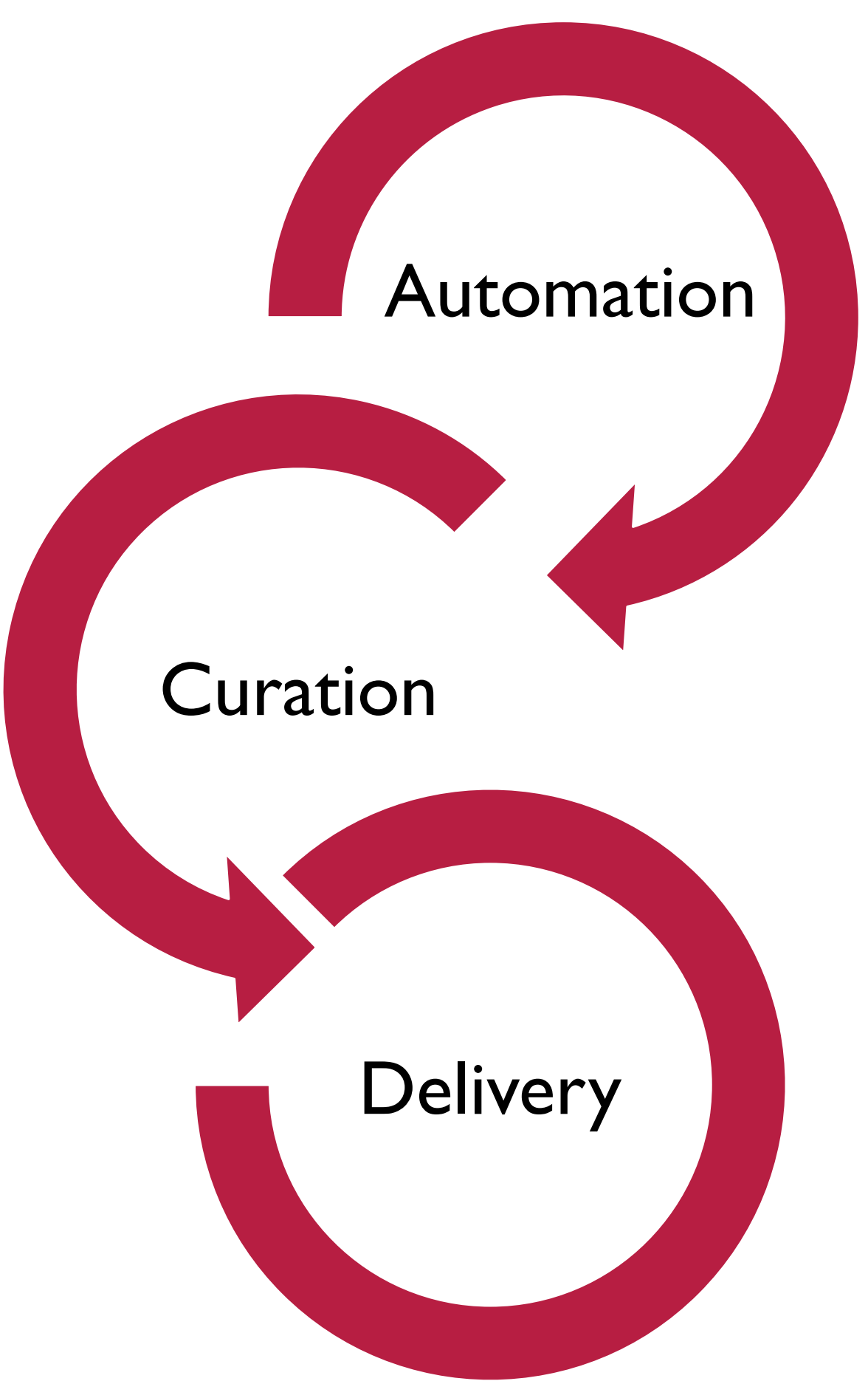
Archana Ahlawat, David Cheng, Fangge Deng, Lucas Fagan, Sherry Feng, Yuanhao Guan, Ethan Holland, Sherry Hu, Drew Learner, Matthew O'Boyle, Ali Soyupak, Caroline Wang, Jason Wang, Connie Wu, Harry Xie, Arthur Zhao, Liuyi Zhu
 Graduate Students: Towqir Aziz, Xinghao Cheng, Mayuresh Kunjir, Junbo Li, Yuhao Wen
 Faculty Leads: Bill Adair (PubPol), Pankaj K. Agarwal (CS/Math), Jun Yang (CS)



Pop-Up Fact-Checking
 Identify claims, check them, and stop false ones in their tracks, right when they are being consumed

Why, and Why Now?

- False claims are often repeated even after being called out
- Many people don't research claims they encounter, and even if they do, they may do so selectively (confirmation bias)
- Thanks to the *ClaimReview* standard at schema.org, we now have a constantly growing database of fact-checks by reputable organizations
- With ubiquity of devices such as smartphones and smart TVs, it is possible to augment what people are hearing/watching with fact-checks in real time

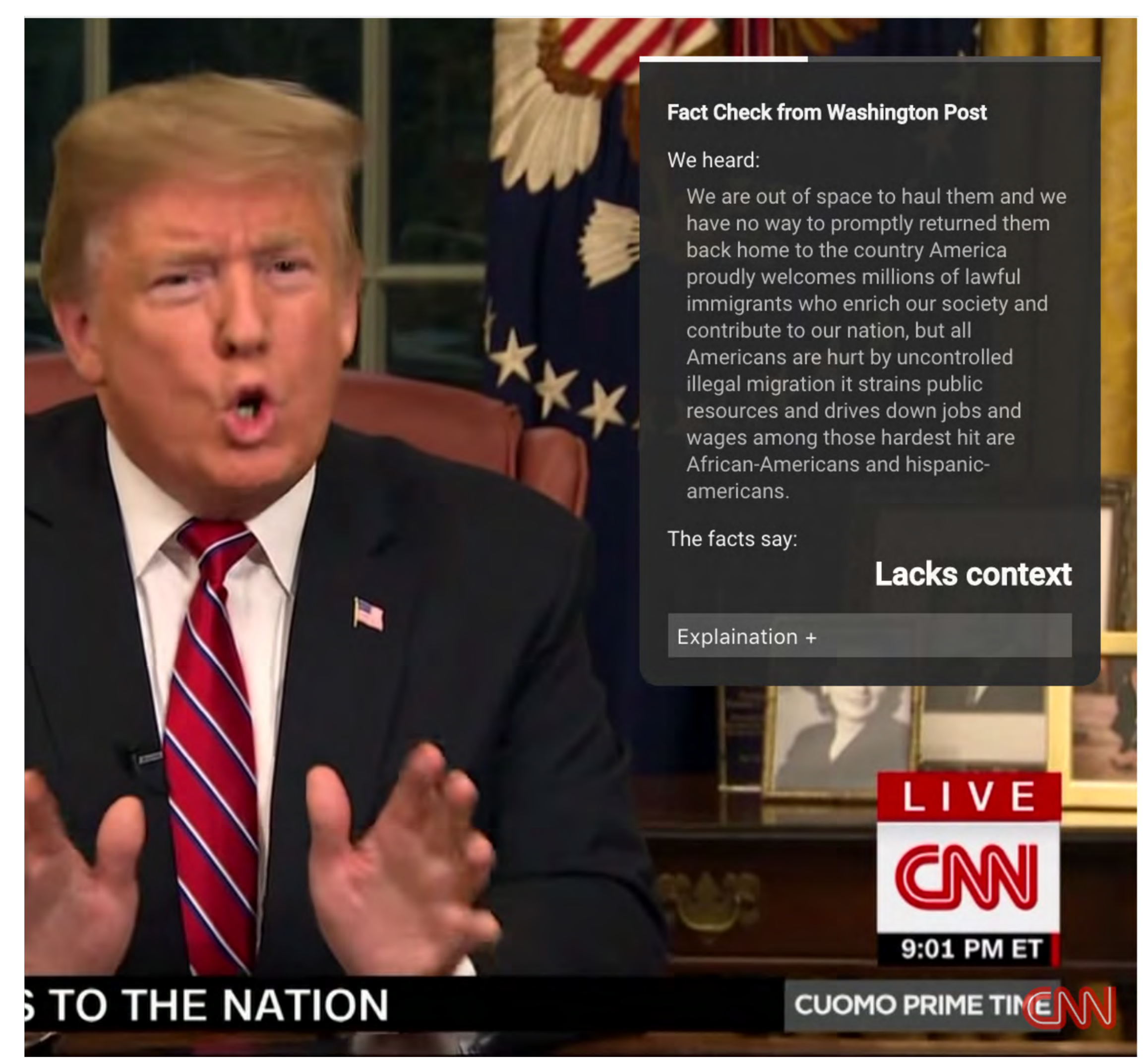


Automation Pipeline

<p>395 sentences</p> <p>Audio Collection We collect live audio from TV channels such as CNN, MSNBC and FOX, from the microphone, and from YouTube (example here shows Trump's 2018 SOTU)</p>	<p>370 well-transcribed sentences</p> <p>Transcription We use Google Cloud Speech-to-Text API</p>	<p>199 candidate claims</p> <p>Checkworthiness Filter We use ClaimBuster to filter out sentences that are not check-worthy factual claims</p>	<p>30 matches</p> <p>Claim Matching We use information retrieval (Lucene indexing) and natural language processing (Transformer-based neural net) techniques to match each claim against a database of fact-checks</p> <p>Share the Facts >19,000 fact-checks</p> <p><i>Currently 65-80% accuracy: need humans in the loop for curation</i></p>
---	--	--	---

Delivering Live Fact Checks

- Goal: deliver a fact-check stream for an A/V stream through multiple platforms: on-screen (TV, web), second-screen (app)
- Challenges: synchronize production/consumption of multiple live streams; challenge/inform but not alienate audience



- < **Chrome Extension**
- Detects if YouTube video being watched has a fact-check stream, and shows it if available
 - Displays fact checks in real time, synchronized with video, even when you pause, rewind, and fast-forward
 - Links to source articles
 - Provides a history sidebar to view earlier fact checks

Examples from Live-Testing the 2019 SOTU

- A good find: *More people are working now than at any time in the history of our country*
- *More people working in U.S.A. today than at any time in our HISTORY. Media barely covers!* (Misleading / FactCheck.org)
 - *More Americans are working today than at any time in our history.* (Lacks Context / FactCheck.org)
- An uncertain match: *Possible nearly 5 million Americans have been lifted off food stamps*
- *Almost 3.9 million Americans have been lifted off food stamps.* (He's close / PolitiFact, but checked in Aug. 2018)
- A poor match: *Members of Congress the state of our Union is strong*
- *Every single strawberry that is served to Members of Congress will have passed through the rough hands of an undocumented immigrant.* (Half True / PolitiFact)

No benefit of hindsight

Ongoing Work

- Better curation tools for fact-checkers
- User study for on-screen delivery of fact-checks
- Smarter claim parsing and IR scoring
- Keep up with advances in NLP; scale up model training and deployment with clusters and GPUs
- Develop better ways to collect more labeled examples
- Extract more metadata and context to help matching
- Speaker identification

> *Get ready for the 2020 Elections!*