

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

**Paper by:**

**Naeemul Hassan      Chengkai Li  
Fatma Arslan      Mark Tremayne**

**Presentation by:**

**Soel Micheletti**







**AP**

The Associated Press

@AP



Foll

Breaking: Two Explosions in the White House and Barack Obama is injured

Reply



Retweet



Favorite



More

2,875

RETWEETS

134

FAVORITES



Source: Google Images



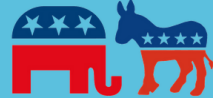


**FACT CHECKING IS VERY IMPORTANT**

# **ClaimBuster: towards end-to-end fact checking**



Enter Your Own Text



U.S. Presidential Debates

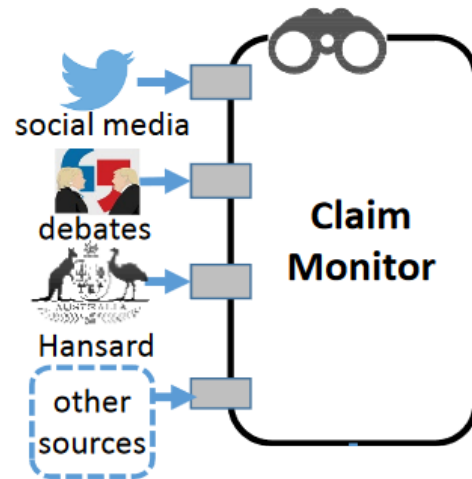
ClaimBuster API

ClaimPortal

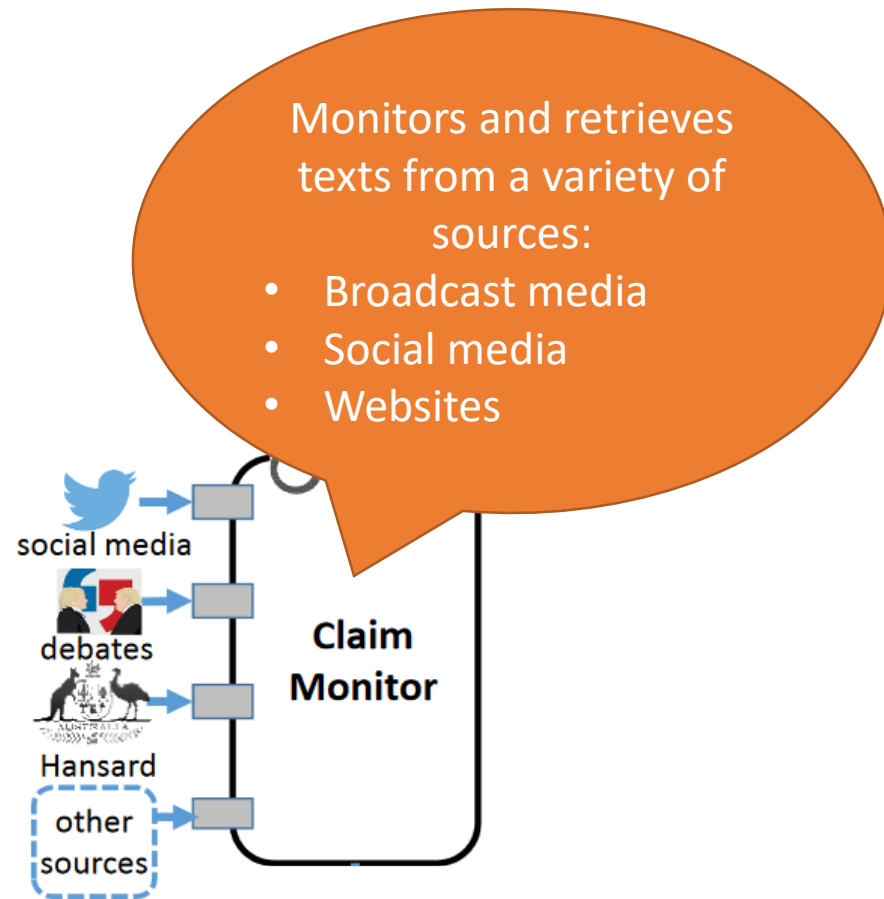
End-to-end Fact-checking (beta)

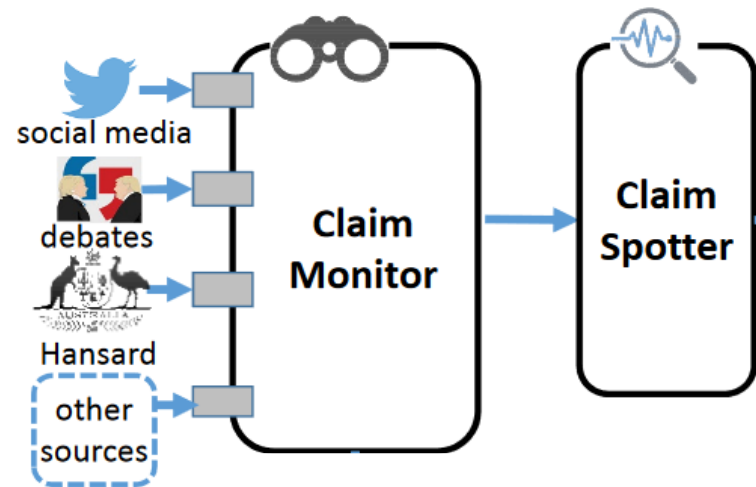
Source: ClaimBuster website

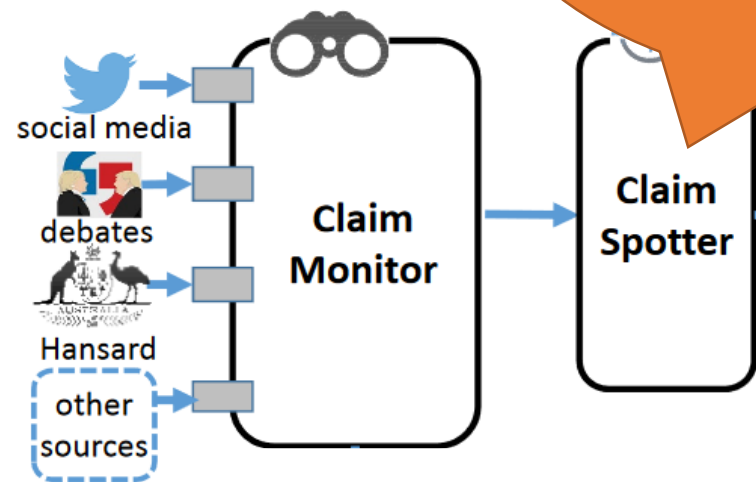
# ClaimBuster: automated end-to-end fact checking



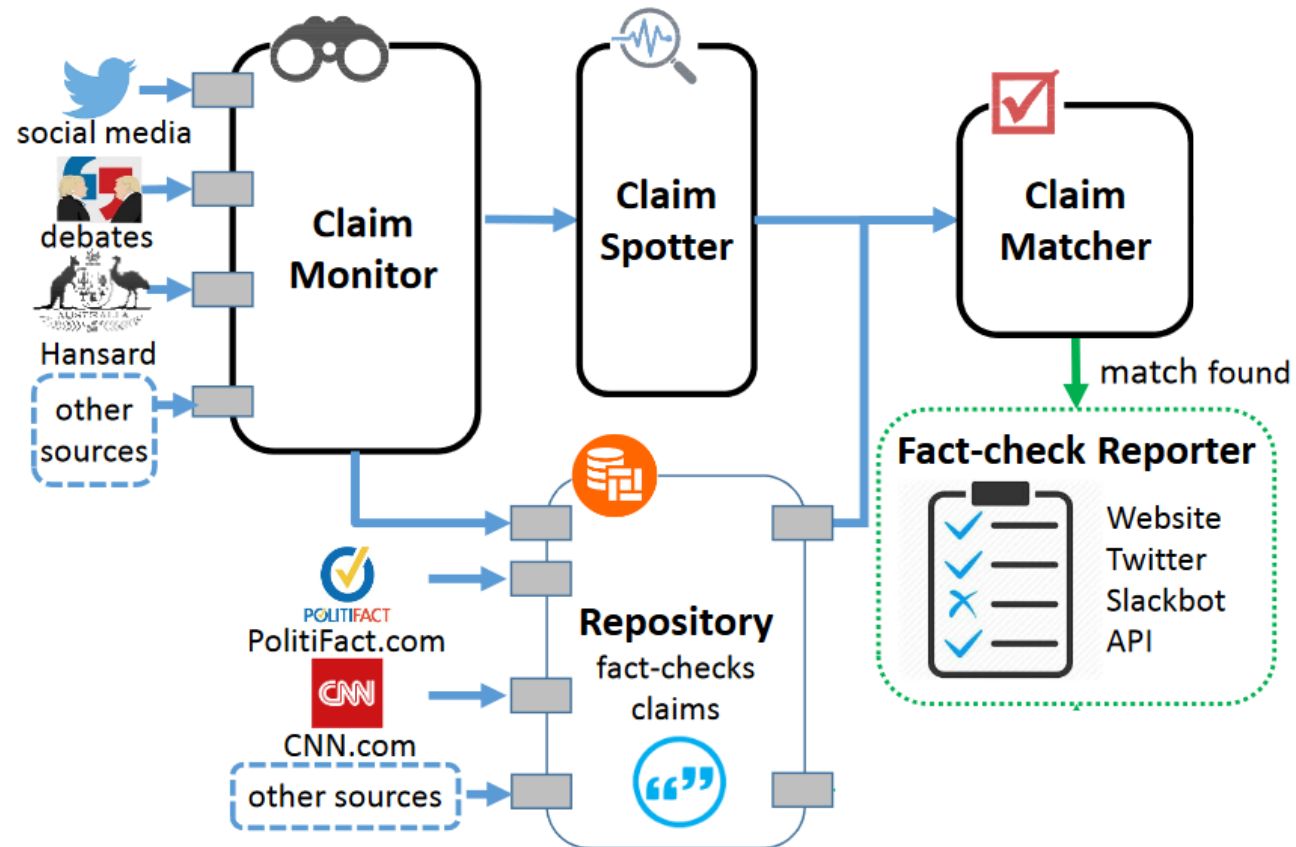




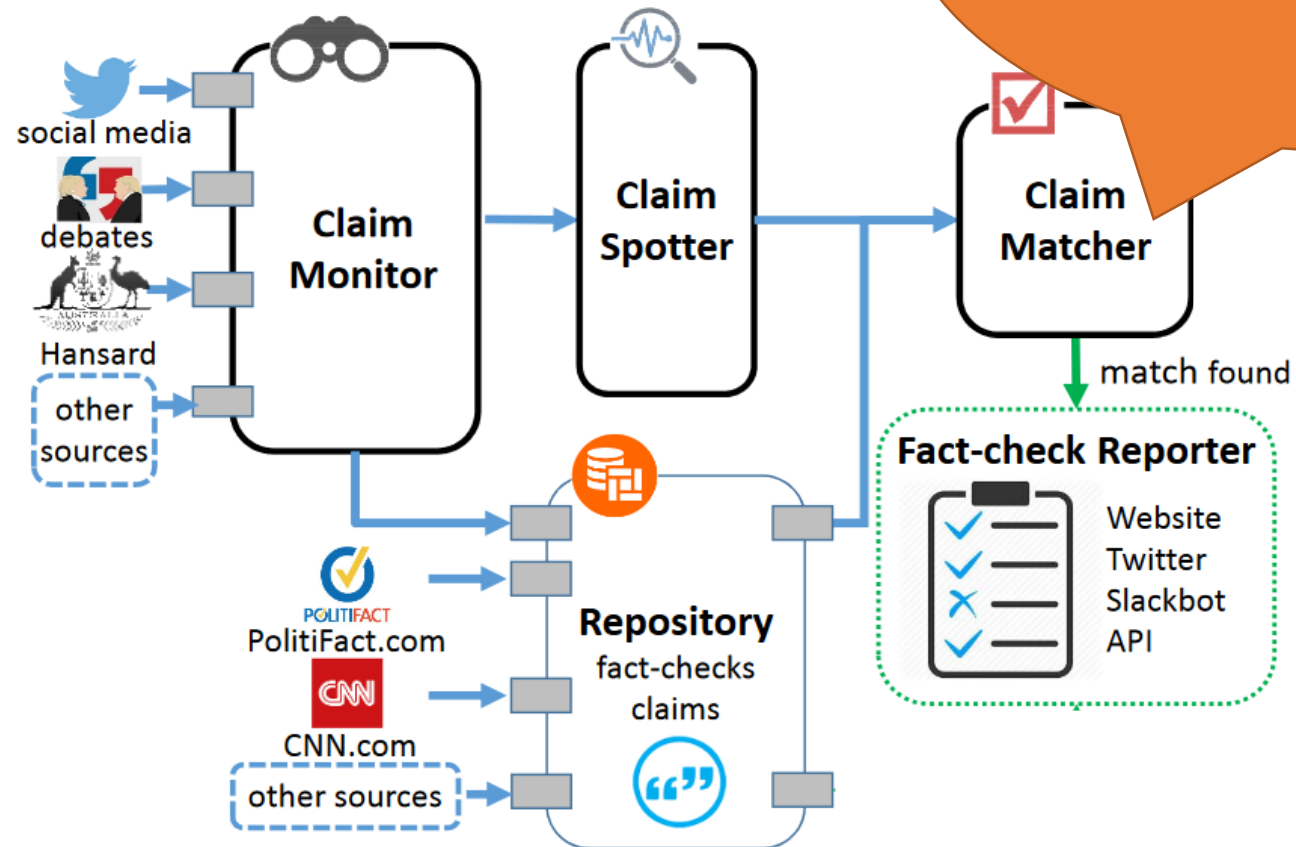




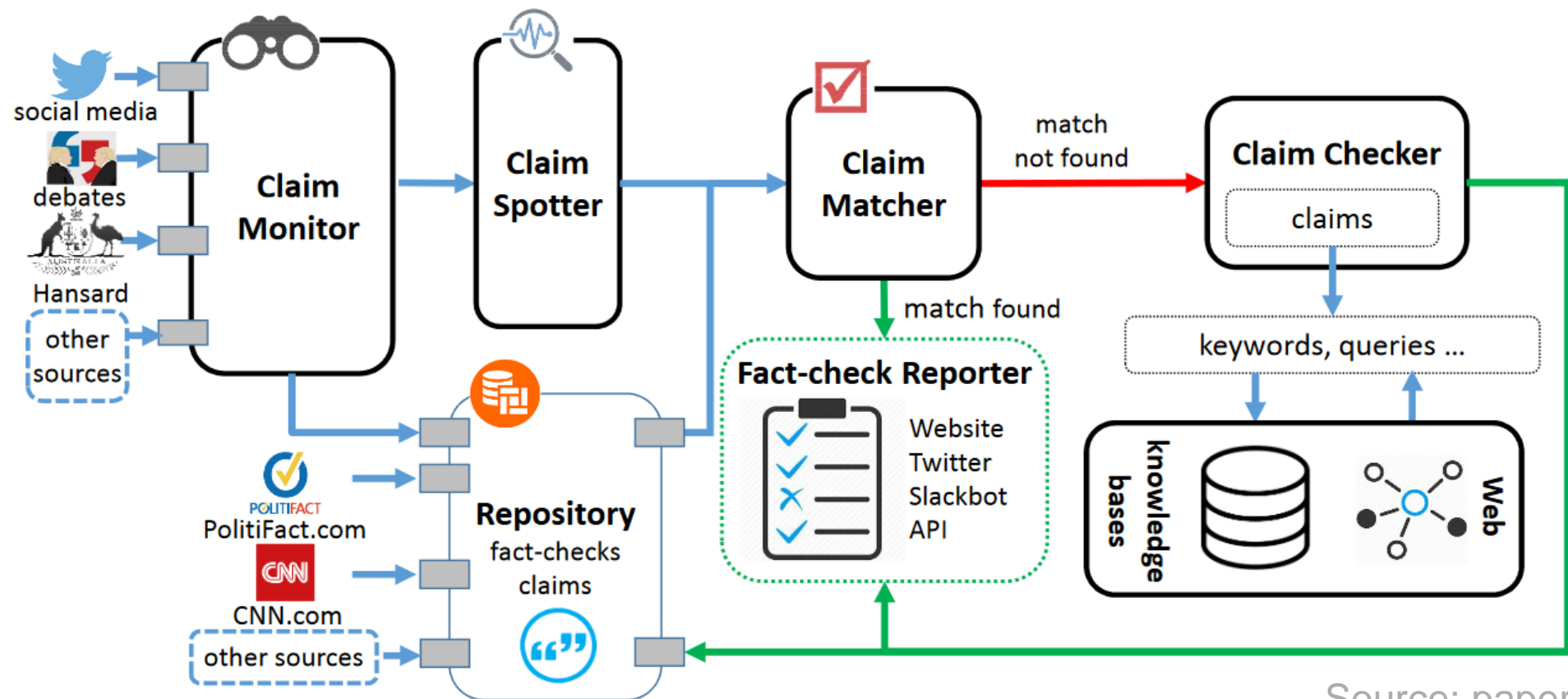
Decides whether a sentence is worth checking.  
MAIN FOCUS OF THE PAPER



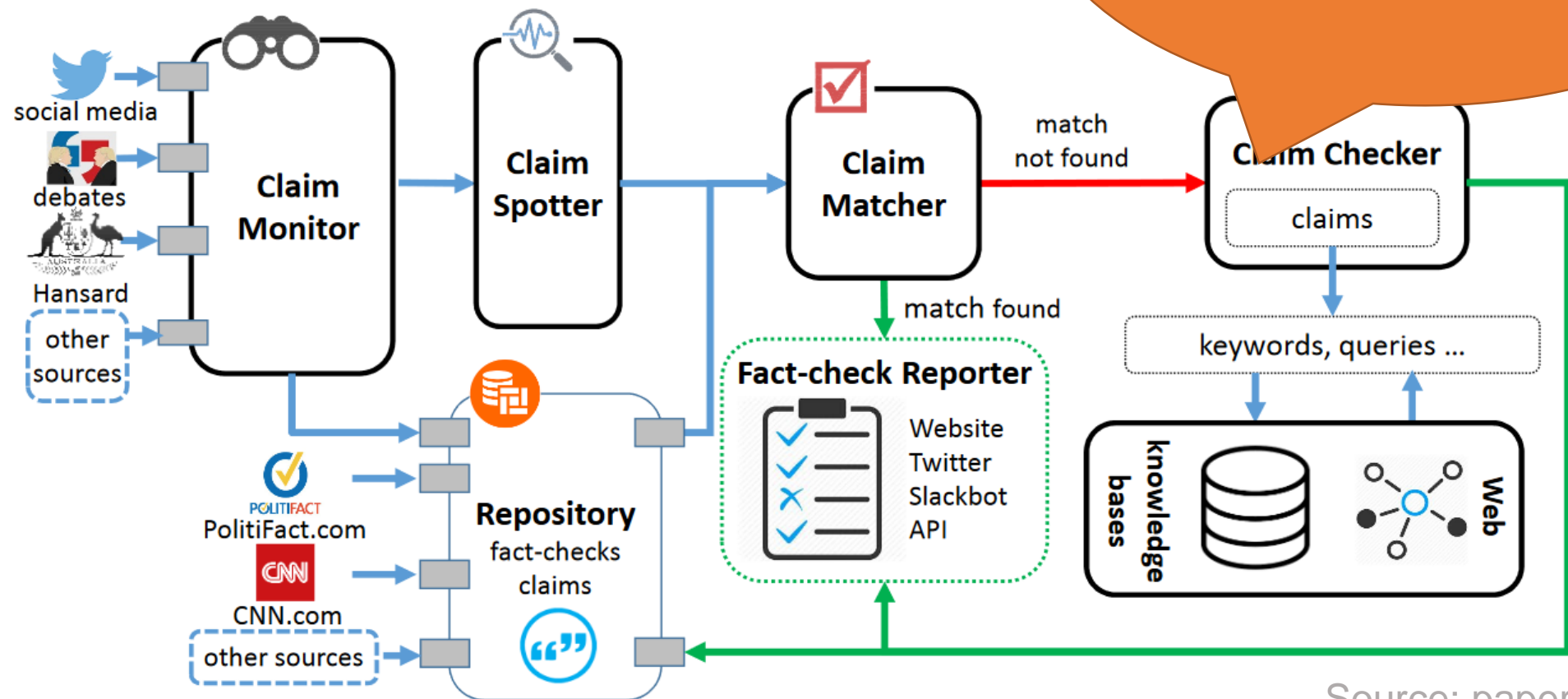




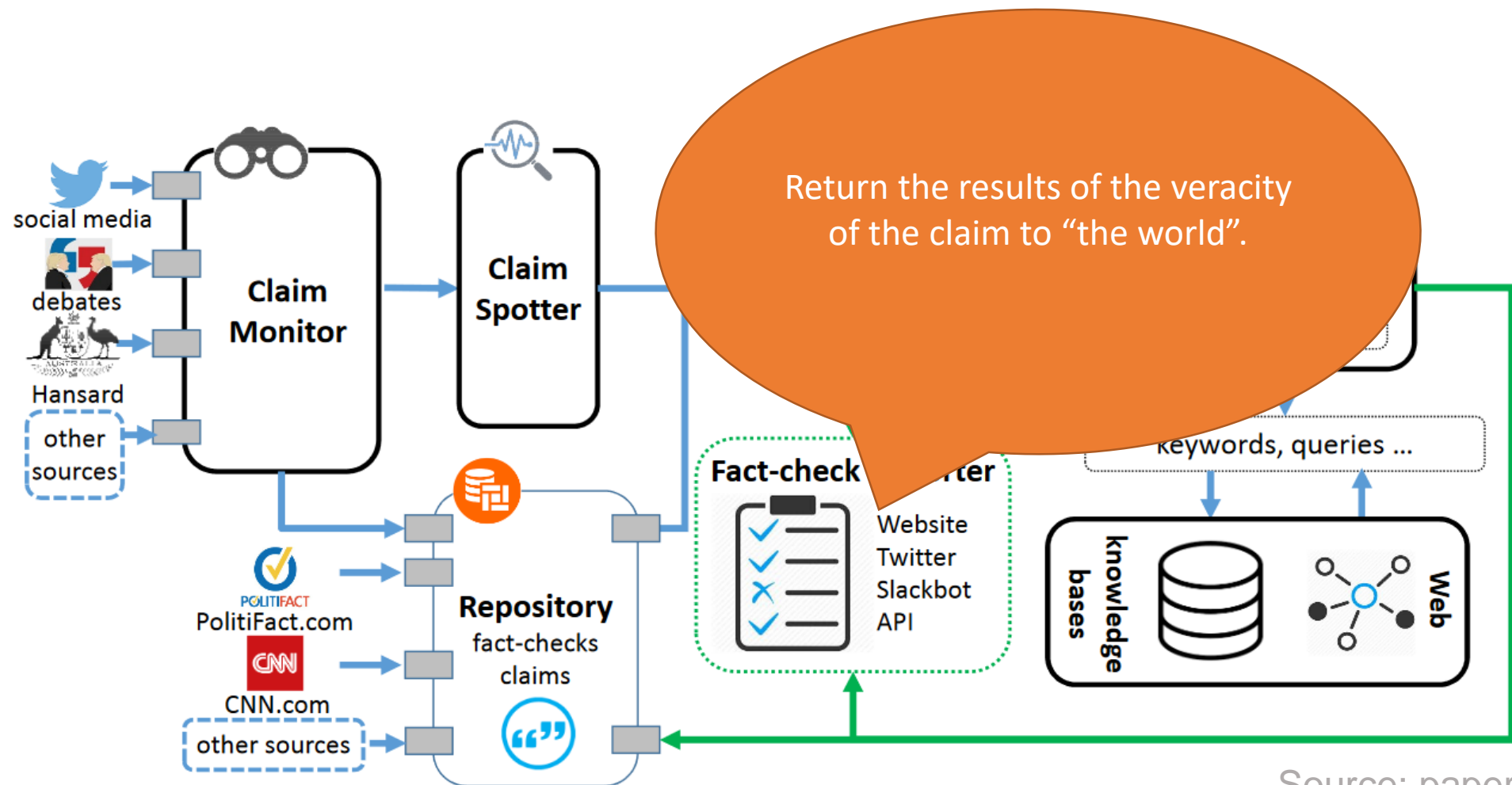
Similar to a “cache”.  
Repository of claims  
previously checked by  
organisations such as  
PolitiFact.



Source: paper



Source: paper







# Claim Spotter

Given a sentence, decide whether it is useful to check it or not

# Claim Spotter

Given a sentence, decide whether it is useful to check it or not

→ A sentence is check-worthy if it contains **important facts!**

# Claim Spotter: find sentences with check-worthy facts

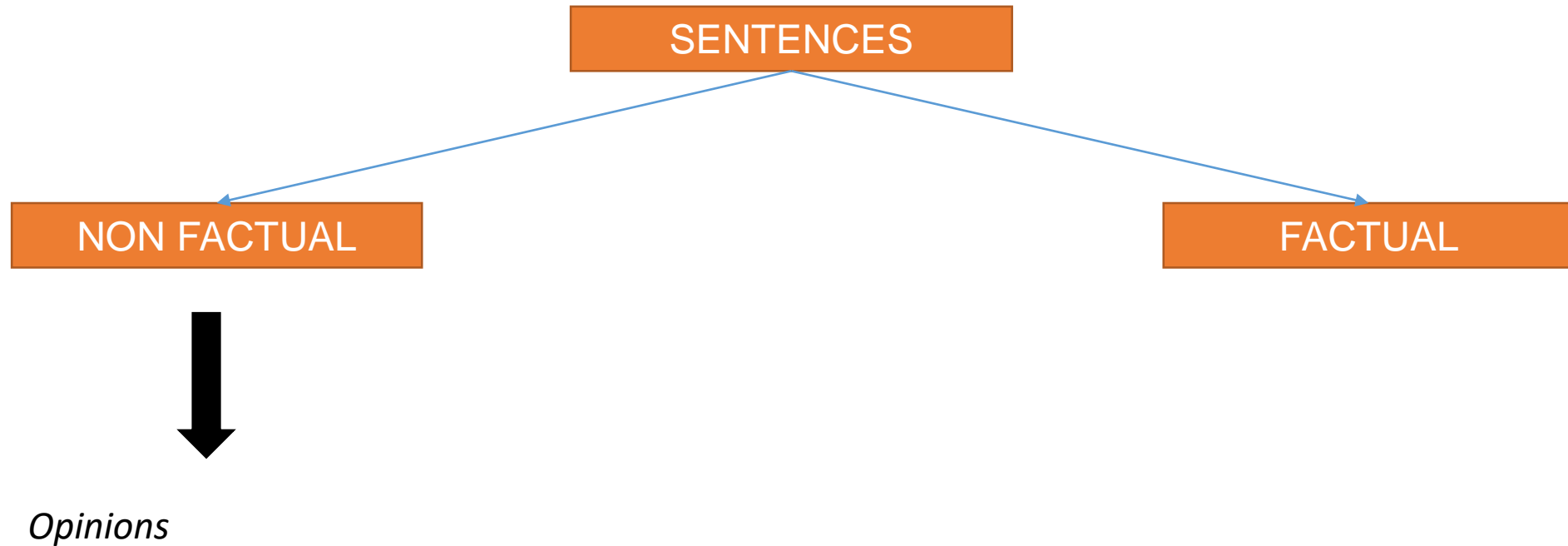
SENTENCES



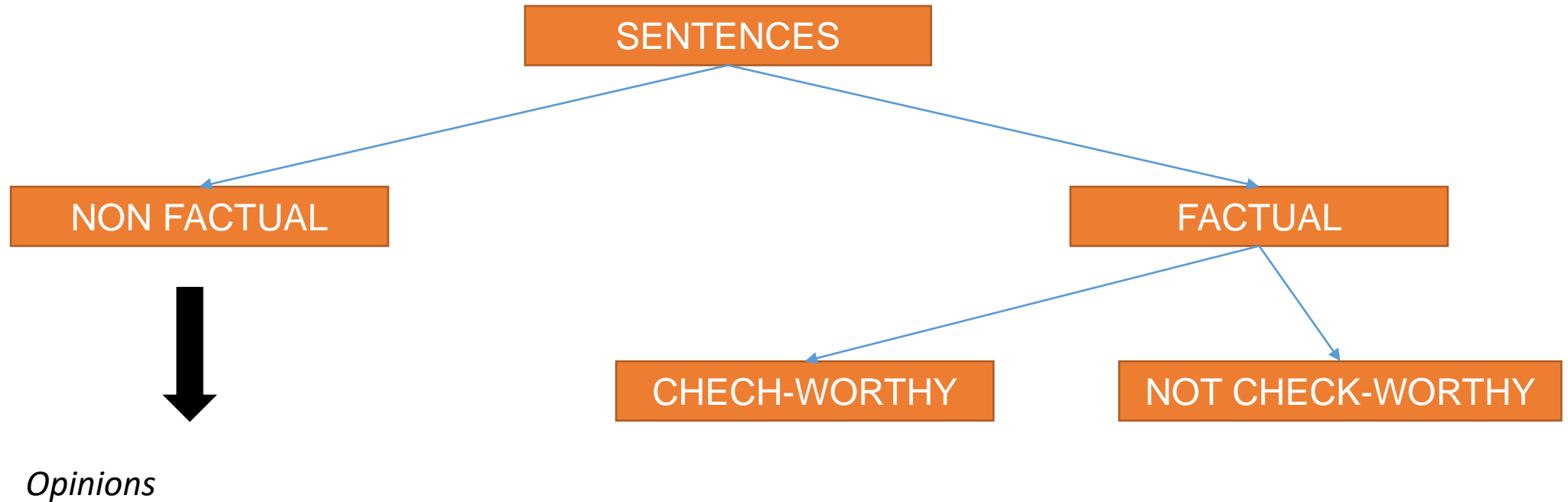
# Claim Spotter: find sentences with check-worthy facts



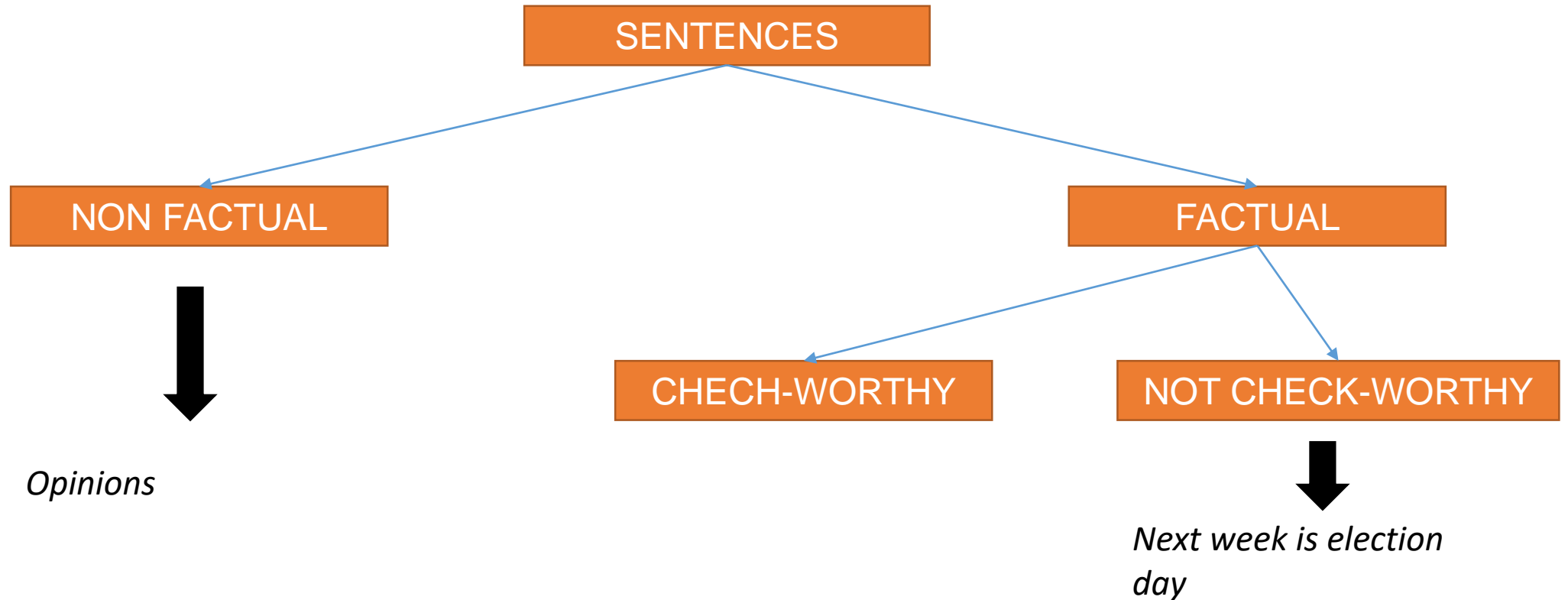
# Claim Spotter: find sentences with check-worthy facts



# Claim Spotter: find sentences with check-worthy facts

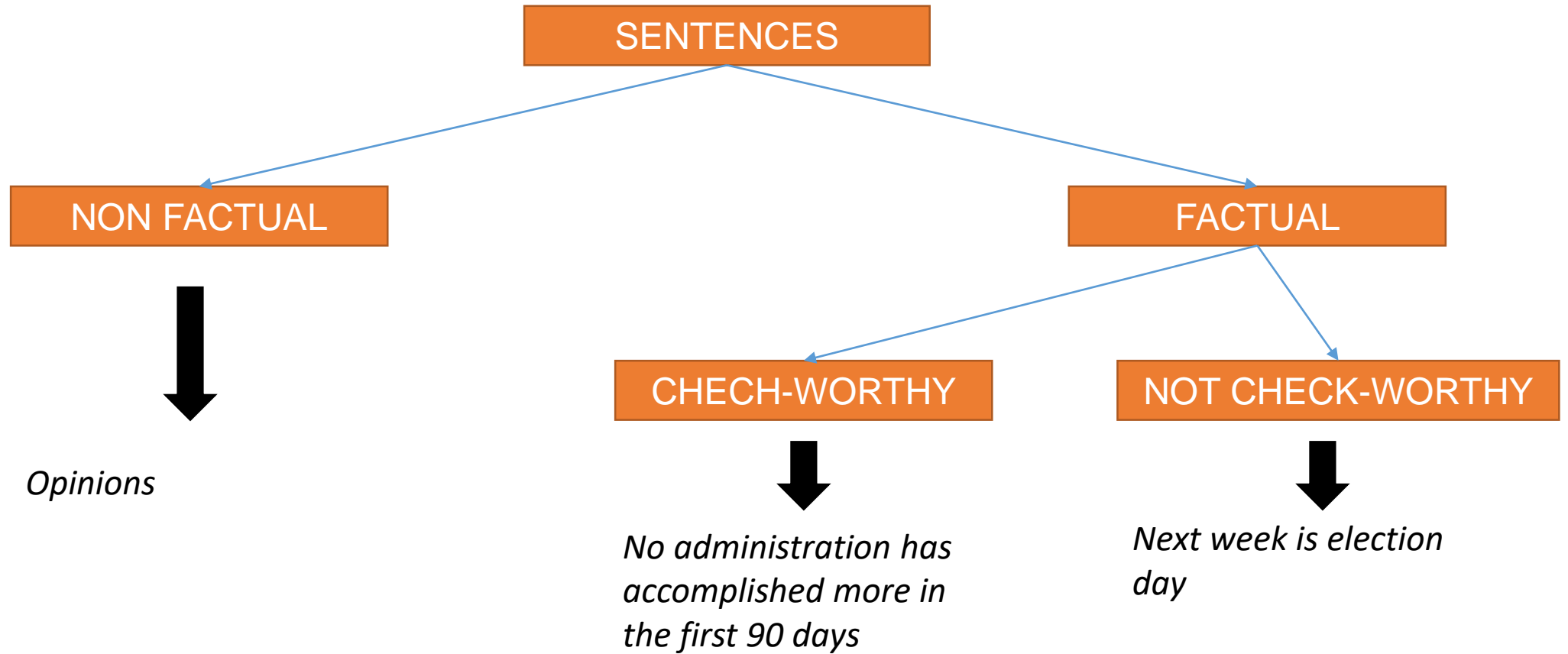


# Claim Spotter: find sentences with check-worthy facts

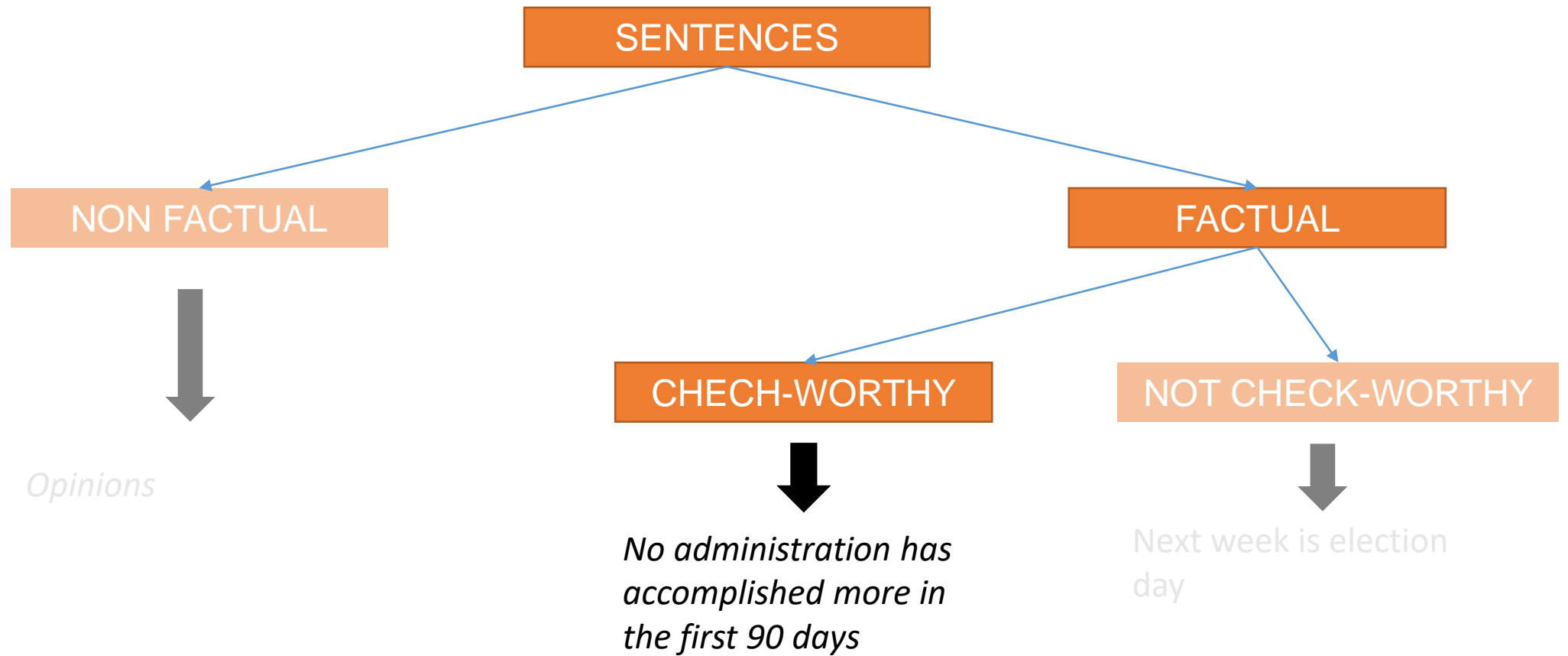




# Claim Spotter: find sentences with check-worthy facts



# GOAL: detect this subset of sentences



# Claim Spotter

Given a sentence, we want to estimate

$\Pr[\textit{class of a statement } x \textit{ is Check-worthy factual sentence}]$

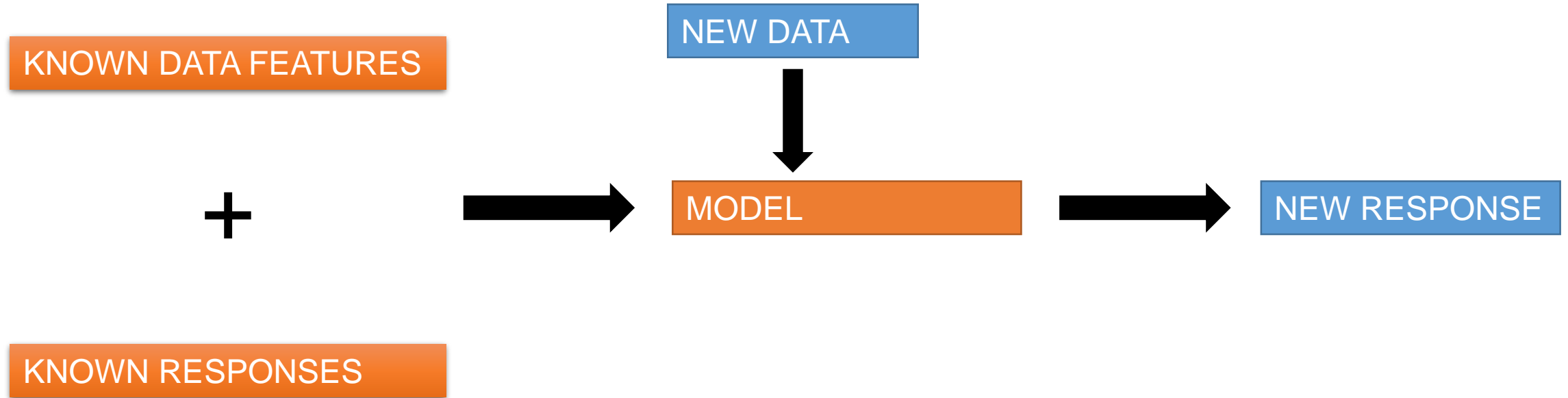
# Claim Spotter

Given a sentence, we want to estimate

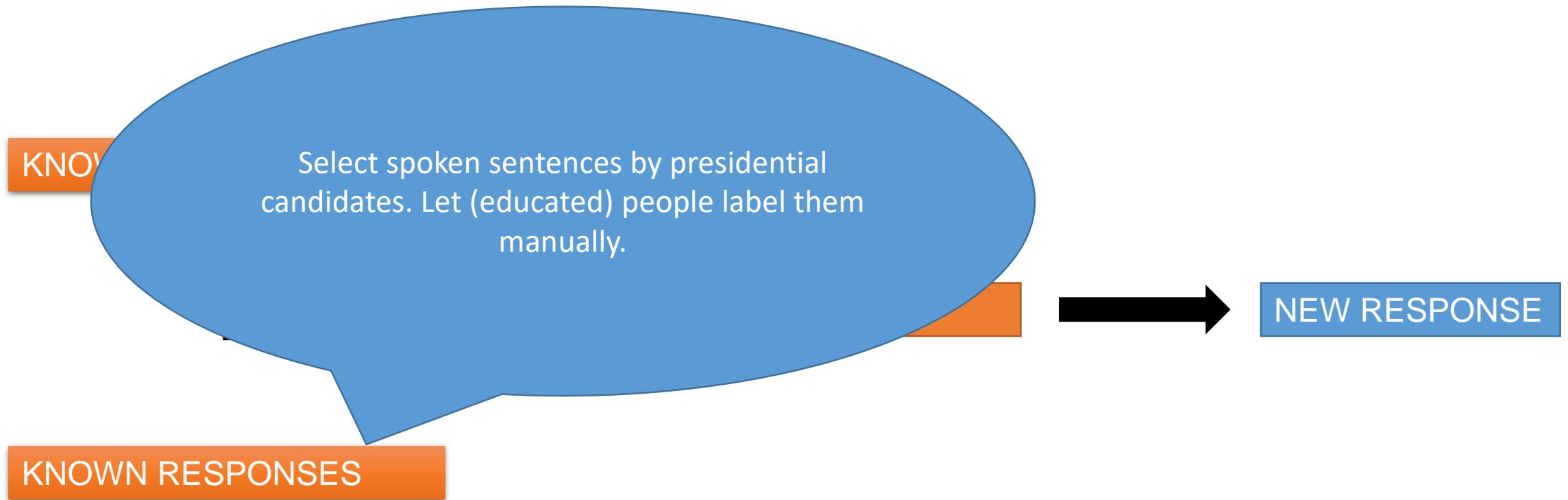
$\Pr[\textit{class of a statement } x \textit{ is Check-worthy factual sentence}]$

DO IT WITH SUPERVISED LEARNING

# Supervised learning



# Supervised learning



# Supervised Learning

Represent most relevant features in a vector.

In this case:

1. Verbs in the past form
2. Cardinal numbers

KNOWN DATA FEATURES

NEW DATA

+



MODEL

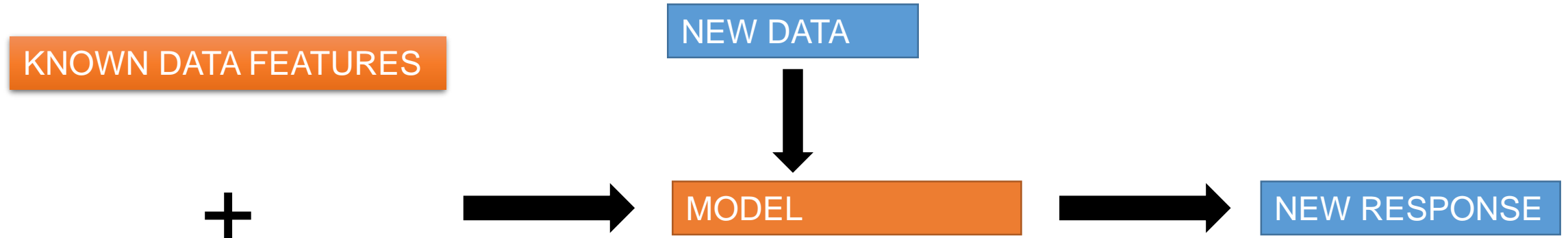


NEW RESPONSE

KNOWN RESPONSES



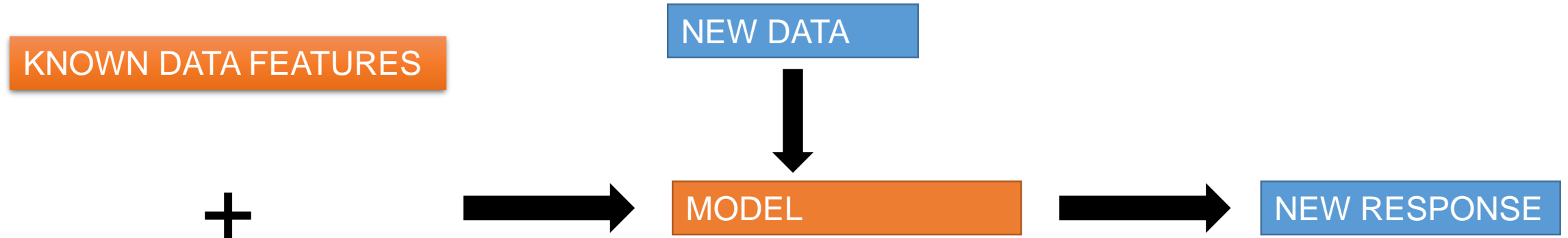
# Supervised learning



MANY MODEL  
POSSIBILITIES:

- CNN
- Support vector machine
- Multinomial Naive Bayes Classifier
- Random forest classifier

# Supervised learning



MANY MODEL  
POSSIBILITIES:

- CNN
- **Support vector machine**
- Multinomial Naive Bayes Classifier
- Random forest classifier

# Claim Spotter

Given a sentence, we want to estimate

$\Pr[\textit{class of a statement } x \textit{ is Check-worthy factual sentence}]$

DO IT WITH SUPERVISED LEARNING

# Claim Spotter

Given a sentence, we want to estimate

$\Pr[\textit{class of a statement } x \textit{ is Check-worthy factual sentence}]$

DO IT WITH SUPERVISED LEARNING

+

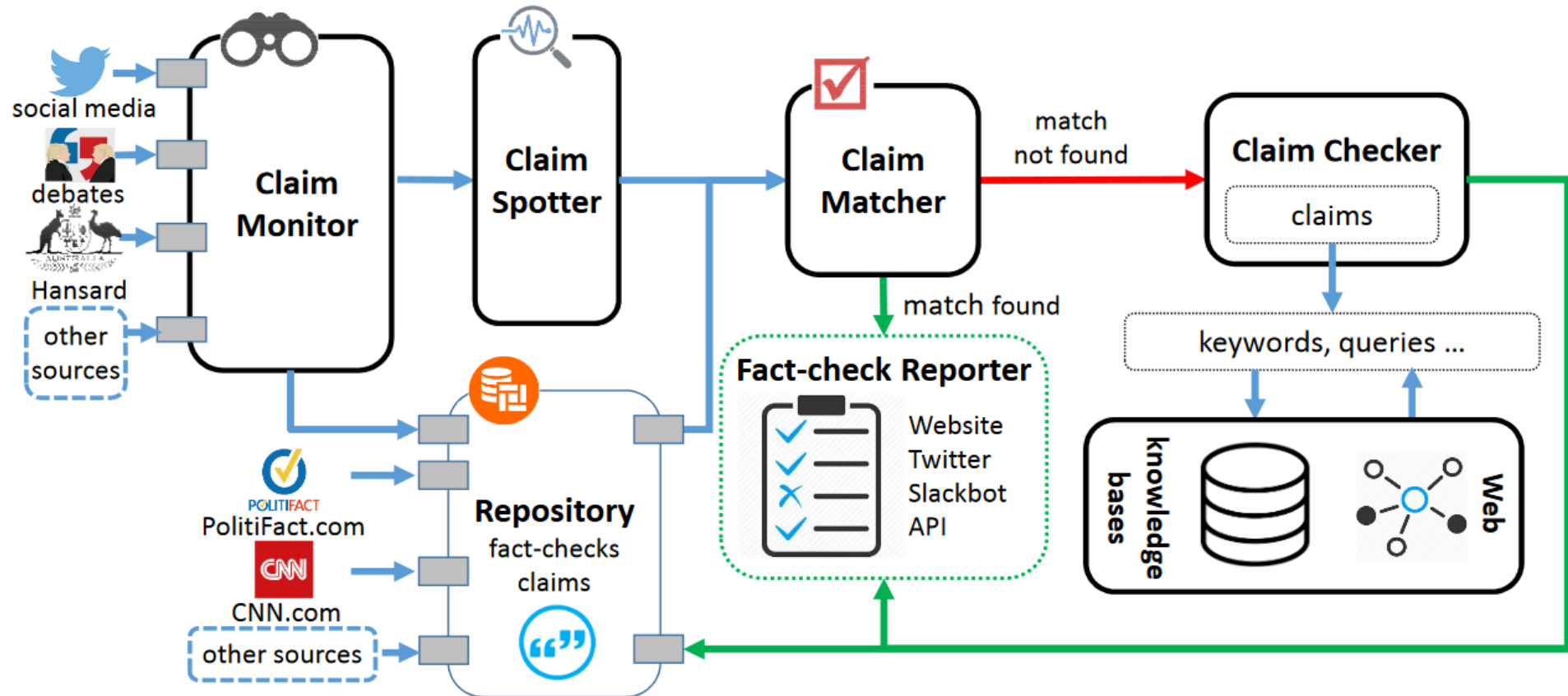
PLATT SCALING

# Results of the claim spotter

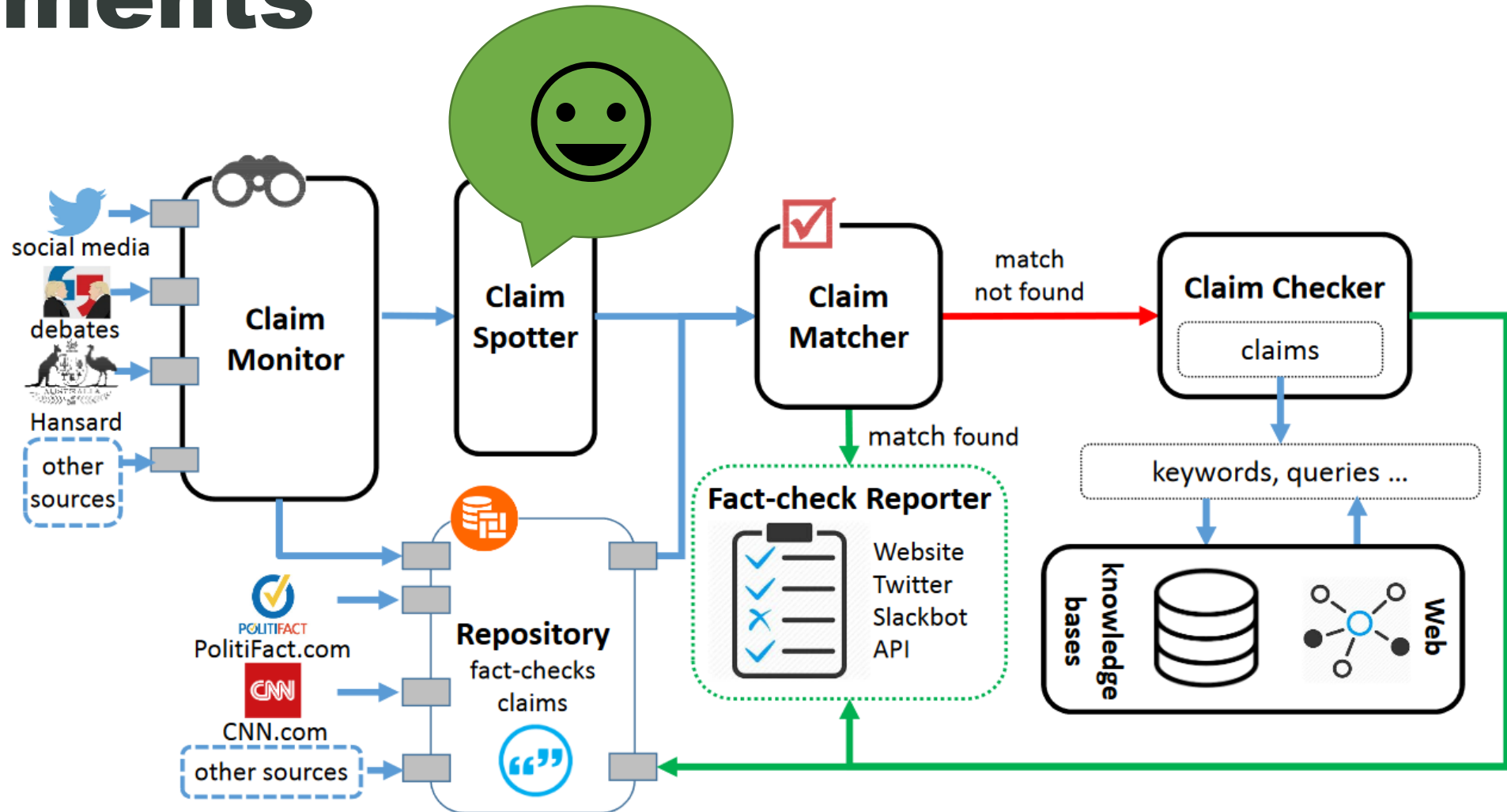
The model is good

**My opinion (plus some  
extension ideas)**

# ClaimBuster: some personal comments

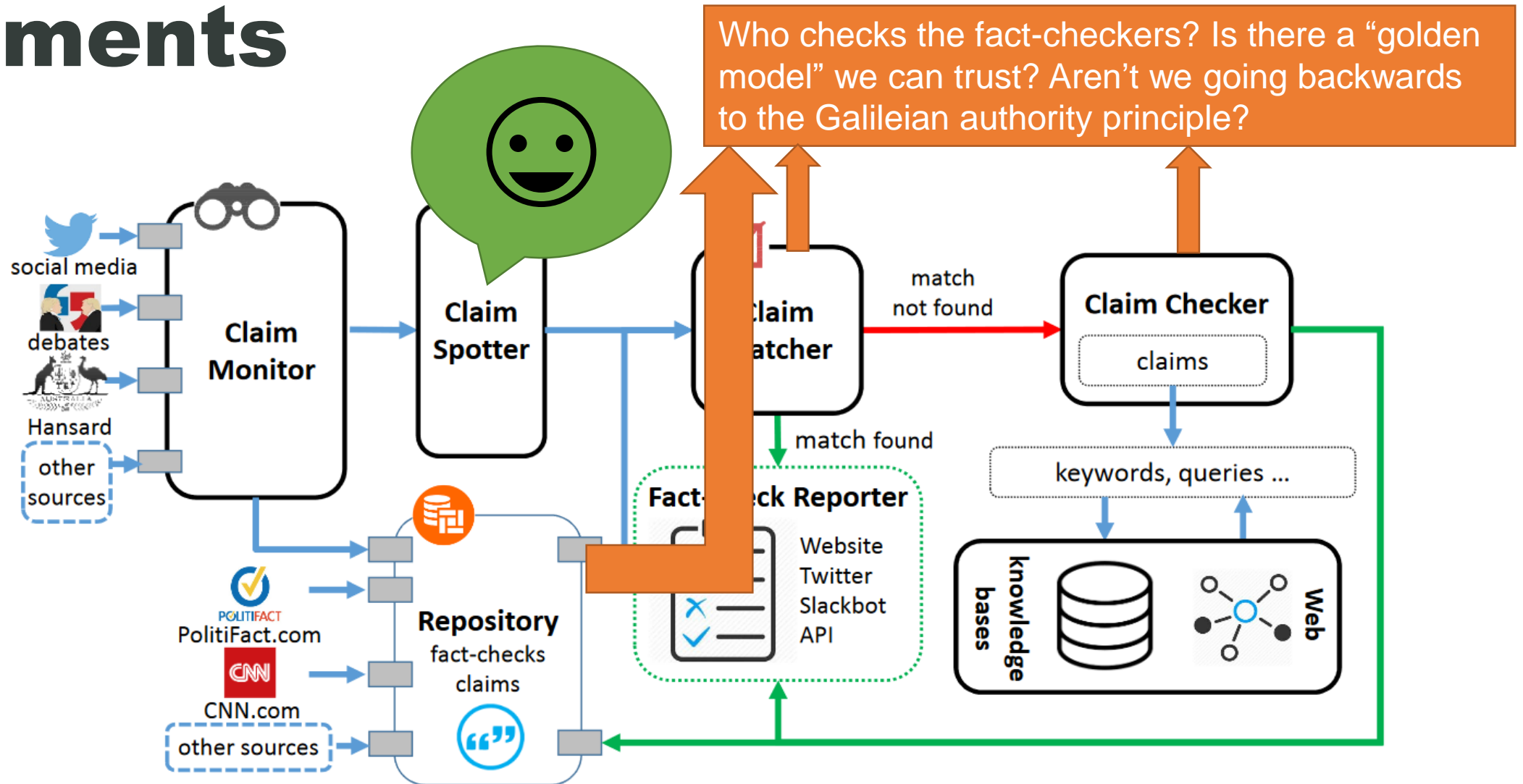


# ClaimBuster: some personal comments

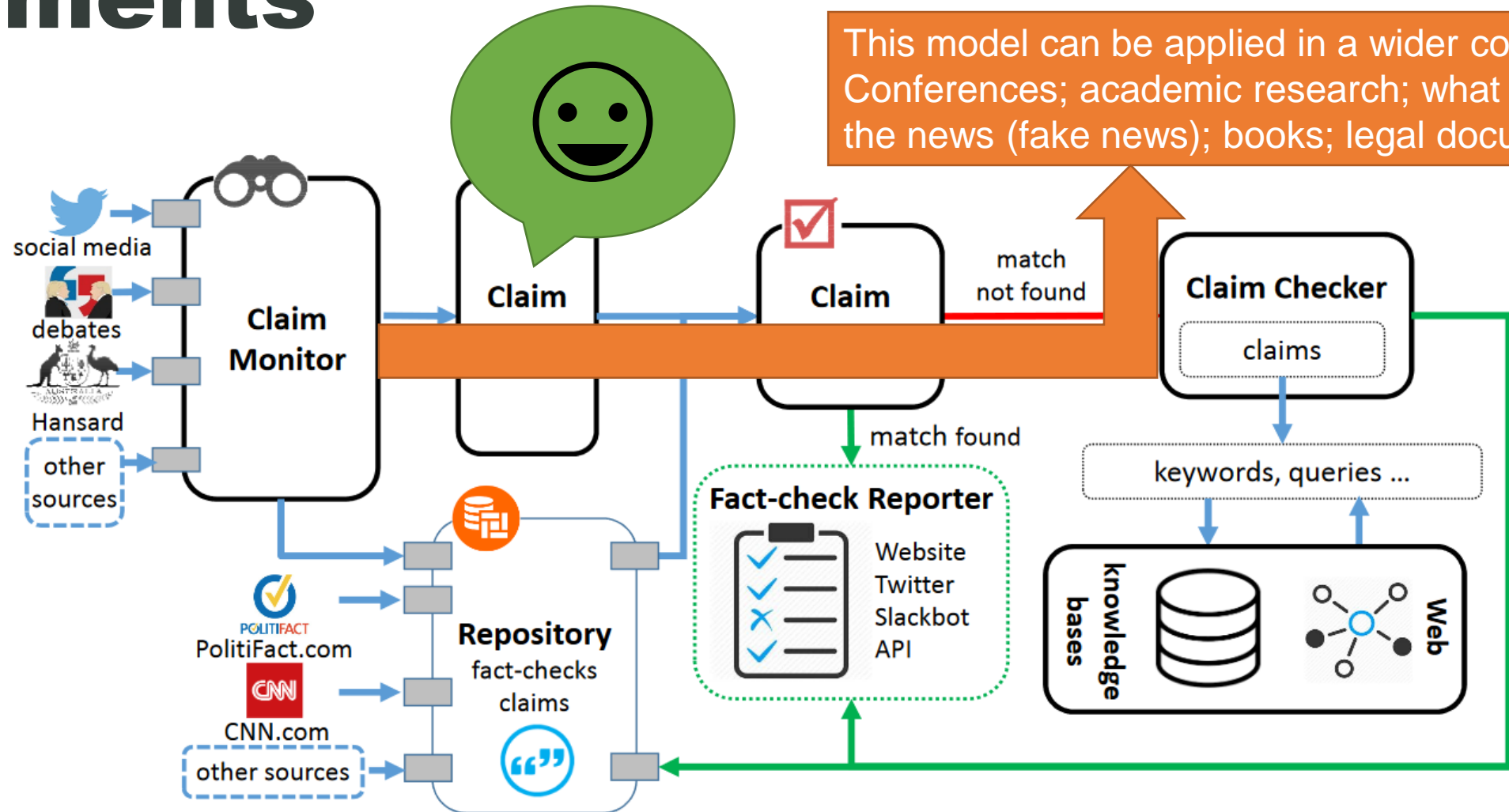




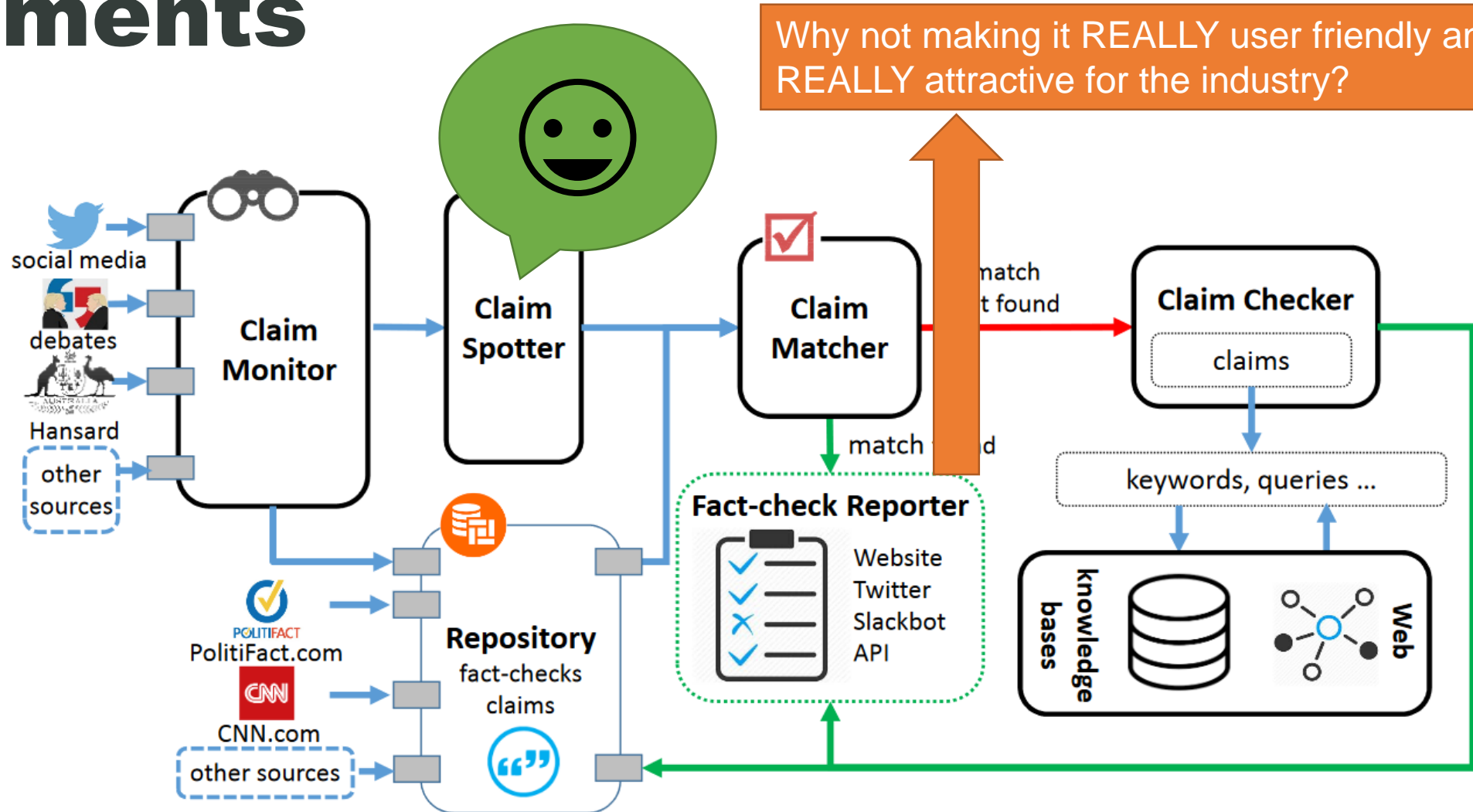
# ClaimBuster: some personal comments



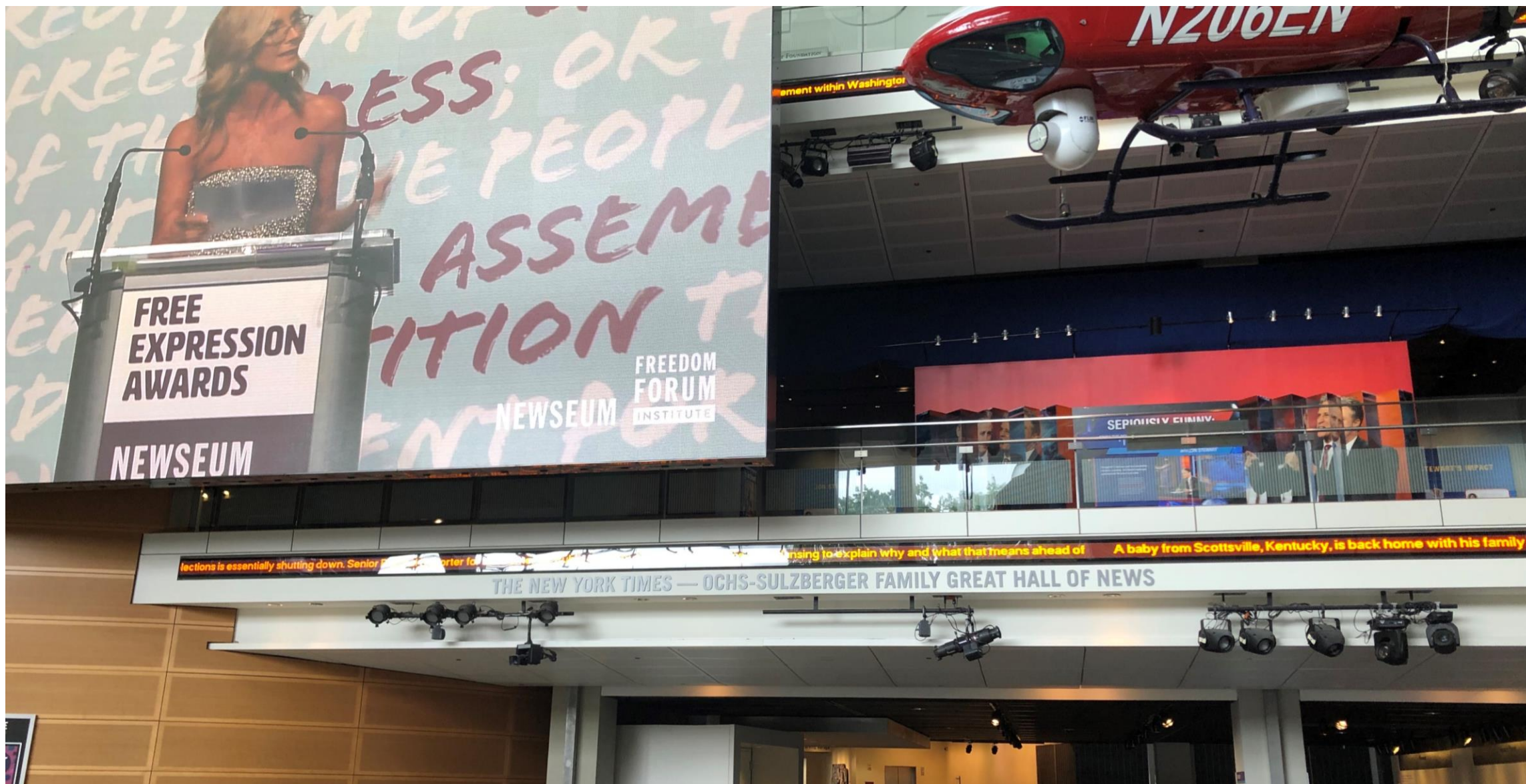
# ClaimBuster: some personal comments



# ClaimBuster: some personal comments








# Backup slides

# What does ClaimBuster offer today?

- Takes claims from different channels (broadcast, social media, websites...)
- Decides what is worth checking
- Checks the most interesting claims

# What does ClaimBuster offer today?

- Takes claims from different channels (broadcast, social media, websites...)
- Decides what is worth checking  Already useful
- Checks the most interesting claims

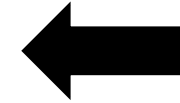
# What does ClaimBuster offer today?

- Takes claims from different channels (broadcast, social media, websites...)
- Decides what is worth checking → Already useful
- Checks the most interesting claims → Big margin for improvement





Enter Your Own Text



U.S. Presidential Debates

ClaimBuster API

ClaimPortal

End-to-end Fact-checking (beta)

Source: ClaimBuster website



Least Check-worthy >=0.1 >=0.2 >=0.3 >=0.4 >=0.5 >=0.6 >=0.7 >=0.8 >=0.9 >=1.0 Most Check-worthy

[Chronological Order](#)[Order by Score](#)

0.48 Donald Trump declared in an interview to the Washington Post that Obama's administration was the best in the History of the United States

[Edit](#)[Submit](#)

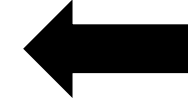
Source: ClaimBuster website



Enter Your Own Text



U.S. Presidential Debates



ClaimBuster API

ClaimPortal

End-to-end Fact-checking (beta)

Source: ClaimBuster website

Least Check-worthy >=0.1 >=0.2 >=0.3 >=0.4 >=0.5 >=0.6 >=0.7 >=0.8 >=0.9 >=1.0 Most Check-worthy

**Savannah Guthrie:** And we love our audience but we would like to ask them to keep their reactions to a minimum and we are not going to hold back making sure that candidates stick to time. So with that business take care—taken care of let's get to it and we are going to start today with Senator Sanders. Good evening to you. You have called for big new government benefits like universal healthcare and free college. In a recent interview you said you suspected that Americans would be quote delighted to pay more taxes for things like that. My question to you is will taxes go up for the middle class in a Sanders administration? And if so how do you sell that to voters?

**Bernard Sanders:** Well, you are quite right. We have a new vision for America and at a time when we have three people in this country owning more wealth than the bottom half of America, while 500,000 people are sleeping out on the streets today we think it is time for change, real change. And by that I mean that healthcare in my view is a human right and we have got to pass a Medicare for all single-payer system. Under that system by the way vast majority of the people in this country will be paying significantly less for healthcare than they are right now. I believe that education is the future for this country and that is why I believe that we must make public colleges and universities tuition-free and eliminate student debt and we do that by placing a tax on Wall Street. Every proposal that I have brought forth is fully paid for.

**Note:** You can also click on a sentence and then click "Fact-check this" to view fact-checking results. The selected sentence will be underlined black.

Claim Checker - Knowledge Bases	Claim Matcher	Claim Checker - Search Engine
Consulting the knowledge bases produced the following results:	We found the following claims which have been professionally fact-checked. Check them out!	We found the following information after processing some search engine results:
<b>Truth Rating</b> Indeterminable <b>Question Asked</b> What is healthcare? <b>Response Recieved</b> noun   the preservation of mental and physical	<b>Truth Rating</b> Mostly True <b>Claim</b> "Today in America, we have more people in jail than any other country on Earth." <b>Speaker</b> Bernie Sanders	<b>No results found</b>

Source: ClaimBuster website



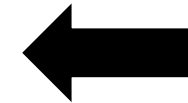
Enter Your Own Text



U.S. Presidential Debates

ClaimBuster API

ClaimPortal



End-to-end Fact-checking (beta)

Source: ClaimBuster website

Includes hashtags:

Claim type:

From these accounts:

Mentioning these accounts:

0

0.5

1

0

0.1

0.2

0.3

0.4

0.5

0.6

0.7

0.8

0.9

1

0.0 - 0.1

0.1 - 0.2

0.2 - 0.3

0.3 - 0.4

0.4 - 0.5

0.5 - 0.6

0.6 - 0.7

0.7 - 0.8

0.8 - 0.9

0.9 - 1.0

Date from:

Date until:

From: 2019-Jan-01

To: 2019-Oct-19

Sort by:

ClaimBuster score (highest first)

1

2

3

4

5

6

7

8

9

10

11

12

13

14

»

**Donald J. Trump**

@realDonaldTrump

After spending more than \$35,000,000 over a two year period, interviewing 500 people, using 18 Trump Hating Angry Democrats & 49 FBI Agents - all culminating in a more than 400 page Report showing NO COLLUSION - why would the Democrats in Congress now need Robert Mueller.....

93.4K
 8:03 PM - May 5, 2019

42.2K people are talking about this

**Donald J. Trump**

@realDonaldTrump

"Nearly one million more blacks and two million more Hispanics are employed than when Barrack Obama left office, and minorities account for more than half of all new jobs created during the Trump Presidency. Unemployment among black women has hovered near 5% for the last six.....

63K
 1:39 PM - Jul 11, 2019

19.5K people are talking about this

Source: ClaimBuster website

ClaimBuster score : 0.8496

**Tweet Text :** Nearly one million more blacks and two million more Hispanics are employed than when Barrack Obama left office, and minorities account for more than half of all new jobs created during the Trump Presidency. Unemployment among black women has hovered near 5% for the last six..... 

### Fact-checked claims:

"The unemployment rate among blacks in Texas is 9.5% -- far more than double white unemployment -- and wages have fallen for black workers in Texas since 2000."

"The unemployment rate among blacks in Texas is 9.5% -- far more than double white unemployment -- and wages have fallen for black workers in Texas since 2000."

"Two million more Latinos are in poverty today than when President Obama took his oath of office less than eight years ago."

that during his eight years as Florida governor, "we created 1.3 million net new jobs -- more jobs created than Texas."

that during his eight years as Florida governor, "we created 1.3 million net new jobs -- more jobs created than Texas."

"In the last 22 months, businesses have created more than 3 million jobs. Last year, they created the most jobs since 2005."

"Our unemployment rate has been higher than 8 percent for more than two-and-a-half years, far above what the Obama Administration promised with the 'stimulus.' "


"Our unemployment rate has been higher than 8 percent for more than two-and-a-half years, far above what the Obama Administration promised with the 'stimulus.' "

"We created 800,000 new jobs, we cut the unemployment rate almost in half and today New York State has more private sector jobs than it has ever had in its history"

Source: ClaimBuster website



# Data labelling

 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.

[Submit](#) [Skip this sentence](#) [Modify My Previous Responses](#)



# Data labelling

- Given a sentence, decide whether it is NFS, UFS or CFS
- **Which sentences do we label in the training set?** 20788 sentences with at least five words from past elections (1960-2012)
- **Who does the manual labeling?** Mostly university students, professors and journalists who are aware of US politics.
- **What about *out of context* sentences?** There is a *more context* button.
- **What if there are spammers?** There are 1032 sentences which were labelled by experts. 10% of the sentences a user classify are from this set and are used to evaluate the participant.

# Data labelling

- **How many times does a sentence have to be labelled in the same way in order to be accepted?** A sentence has to be labelled in the same way by a reasonable number of top-quality participants and most of the people should have labelled it in the same way.

# Feature extraction

- 6615 features in total. Those are too much (overfitting, computational effort too big)
- Examples of features: length of the sentence; number of words; part-of-the speech (verbs in past form, cardinal numbers, percentages, ...), sentiment score (found with AlchemyAPI)
- It is important to detect the best discriminating factors.
- **How to extract best discriminating factors?** Random forest classifier with GINI index, to measure the importance of features in constructing each decision tree.

# Feature extraction

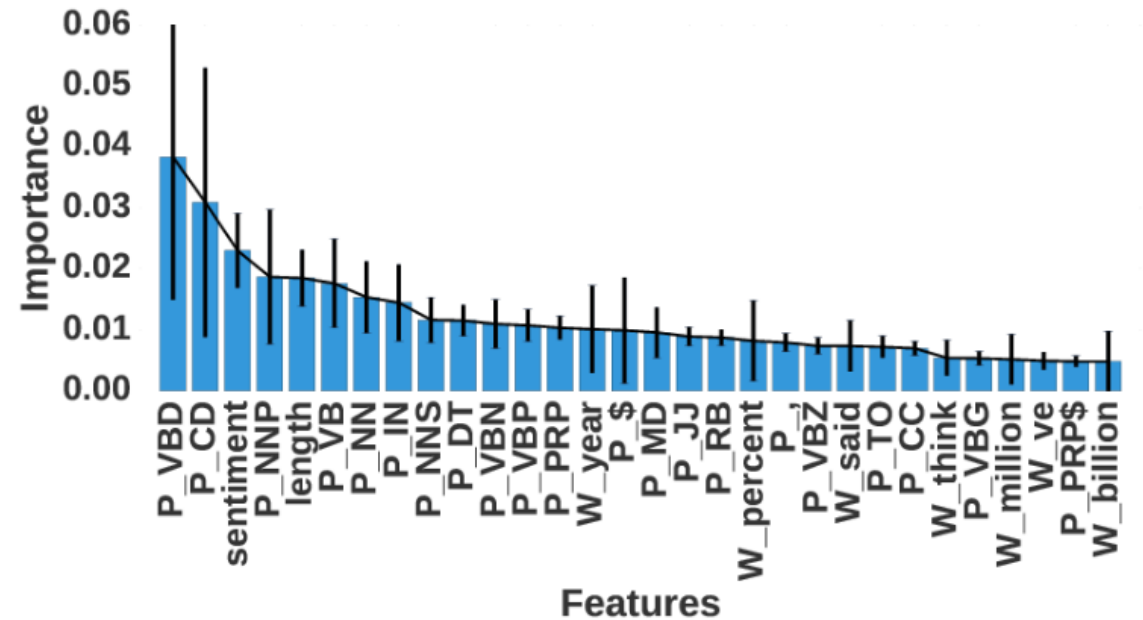


Figure 3: Feature importance