Applied AI & The Fake News Problem

Mike Tamir Ph.D.
mike.tamir@berkeley.edu
Berkeley
UNIVERSITY OF CALIFORNIA

Agenda

1. The Fake news Problem
    a. Solving the right problem
2. Solving the Problem
    a. Deep Learning Representation: Latest Techniques
    b. How it works: Applying the Algorithm
3. The Algorithm in Action

The Fake News Problem

Fake News: The Problem

- Large players are having difficulty preventing malicious use of their services to propagate “fake news”
- Why?
  - **Solving the wrong problem!**
  - For Bigger Social Media and Search enterprises:

    Stopping Fake News amounts to Censorship
- But Censorship is very serious:
  - Requires **Truth Detection**
    - Not just about the article
    - About **article ↔ world** correspondance

Fake News: The *Right* Problem

Image Source

Strong scientific evidence* that

Increased emotional content

negatively impacts cognitive function

* Studies in the field of cognitive control emotion interactions in healthy adults.

Solving “Fake News” means...

Enable readers to detect when the article’s goal is

credible information sharing
vs.
emotional manipulation.

Image Source

Solving “Fake News” means...

Enable **cognitive engagement**

Alert and prevent
manipulated emotional reaction

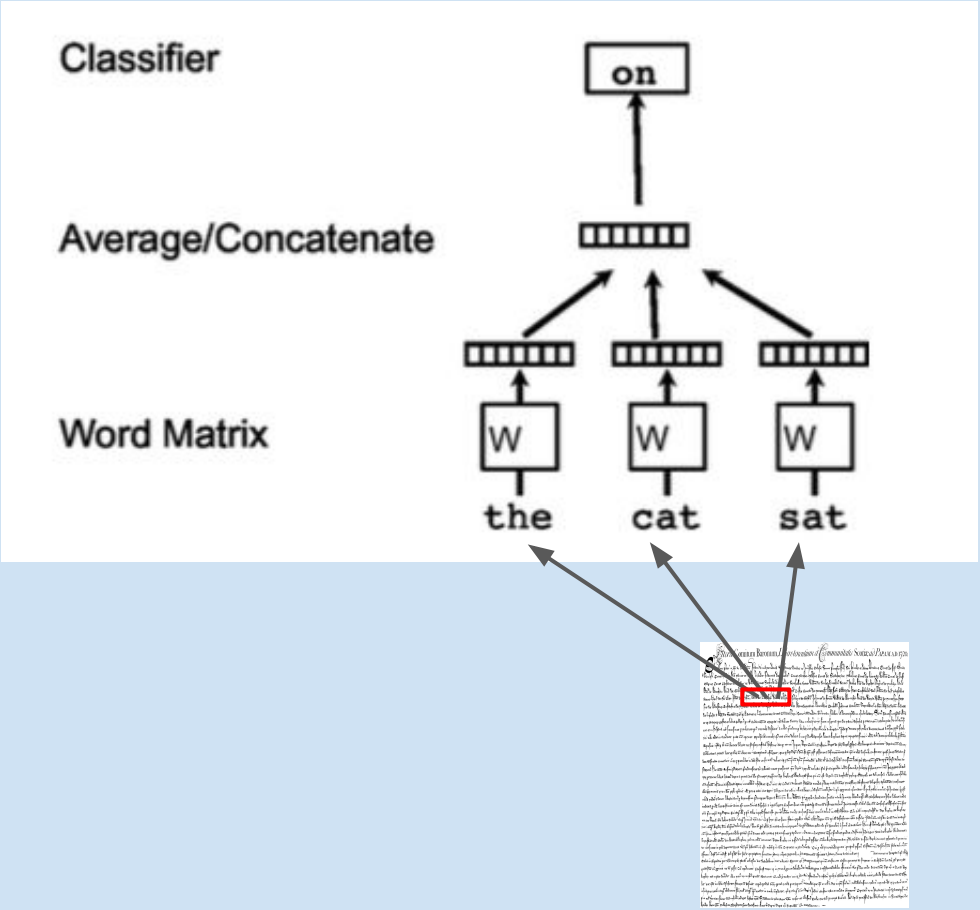
Image Source

Solving the Problem: Deep Learning Techniques

Word2Vec: Algorithm

N-gram ANN classifier:

1. Project the “context” h_j (w_{j-n+1} to w_{j-1})
2. Soft-max predictor for output layer
3. Use BackProp algorithm to execute gradient descent.



Word2Vec: Semantics in the Geometry

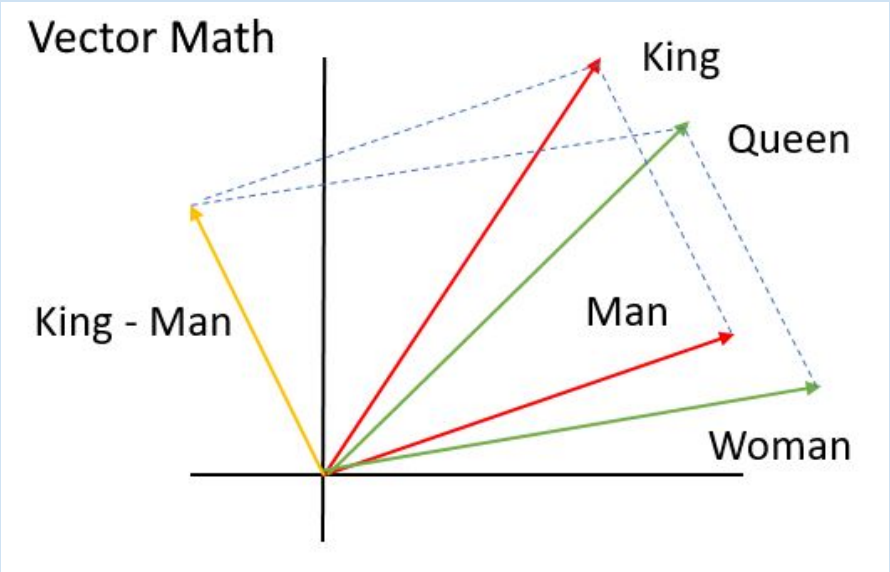


Image from Mathworks

- $|king\rangle - |man\rangle + |woman\rangle = |\boldsymbol{queen}\rangle$
- $|England\rangle - |London\rangle + |Paris\rangle = |\boldsymbol{France}\rangle$
- $|Microsoft\rangle - |Nadella\rangle + |Zuckerberg\rangle = |\boldsymbol{Facebook}\rangle$
- ...

** While W2V was originally achieved with (shallow) Neural Networks, it has been shown that PMI based matrix factorization will generate similar results.*

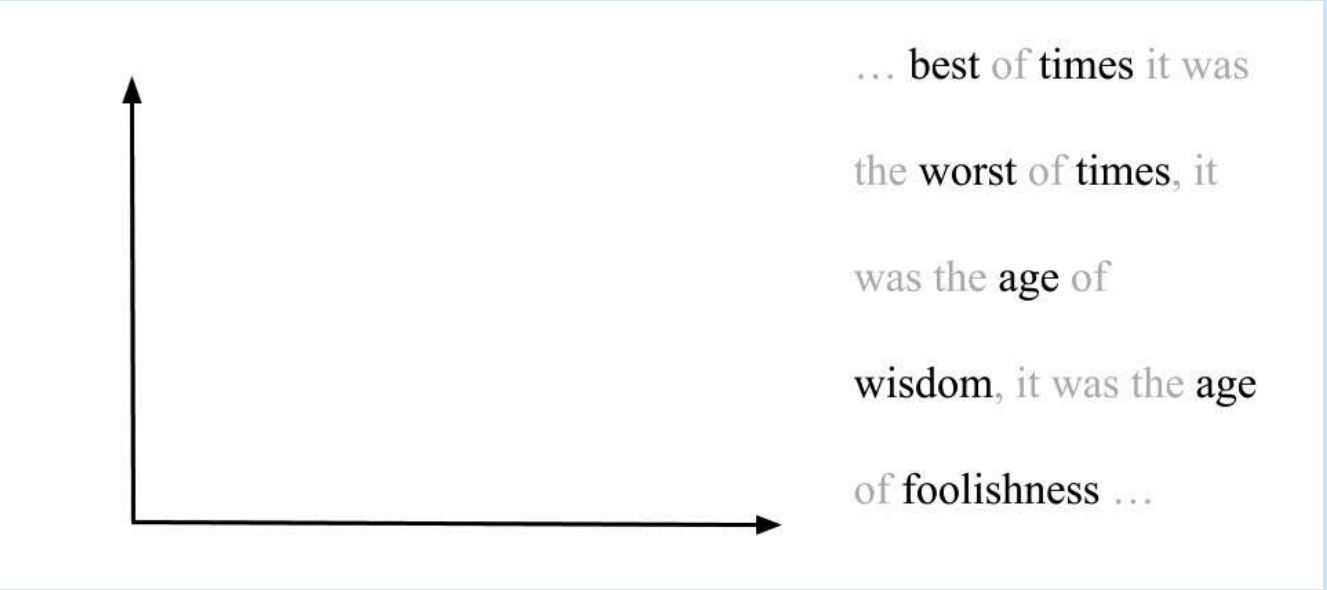
* Levy and Goldberg (NIPS 2014)

Popular Word Embedding Options:

- `Word2Vec`
  - Mikolov et al (2013) Efficient Estimation of Word Representations in Vector Space
  - Google's Gensim Implementation
- `Glove`
  - Pennington et al (2014) - GloVe: Global Vectors for Word Representation
  - Stanford NLP implementation
- `FastText`
  - Joulin et al (2016) - Bag of Tricks for Efficient Text Classification
  - FacebookResearch implementation

Naive Doc2Vec

Take the L2 normalization of the vector sum of terms in the document creates a **random walk** with **bias** process in the w2v space.

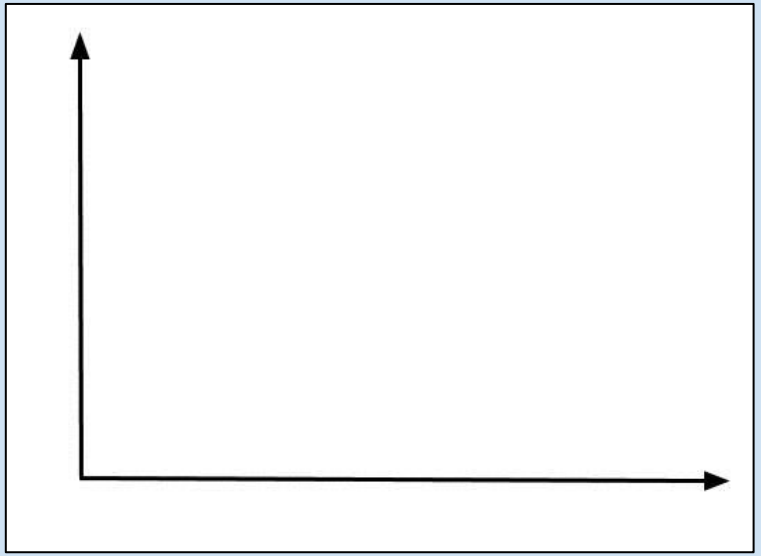


Naive Doc2Vec

ND2V: Easy to implement and captures large scale aggregate signal.

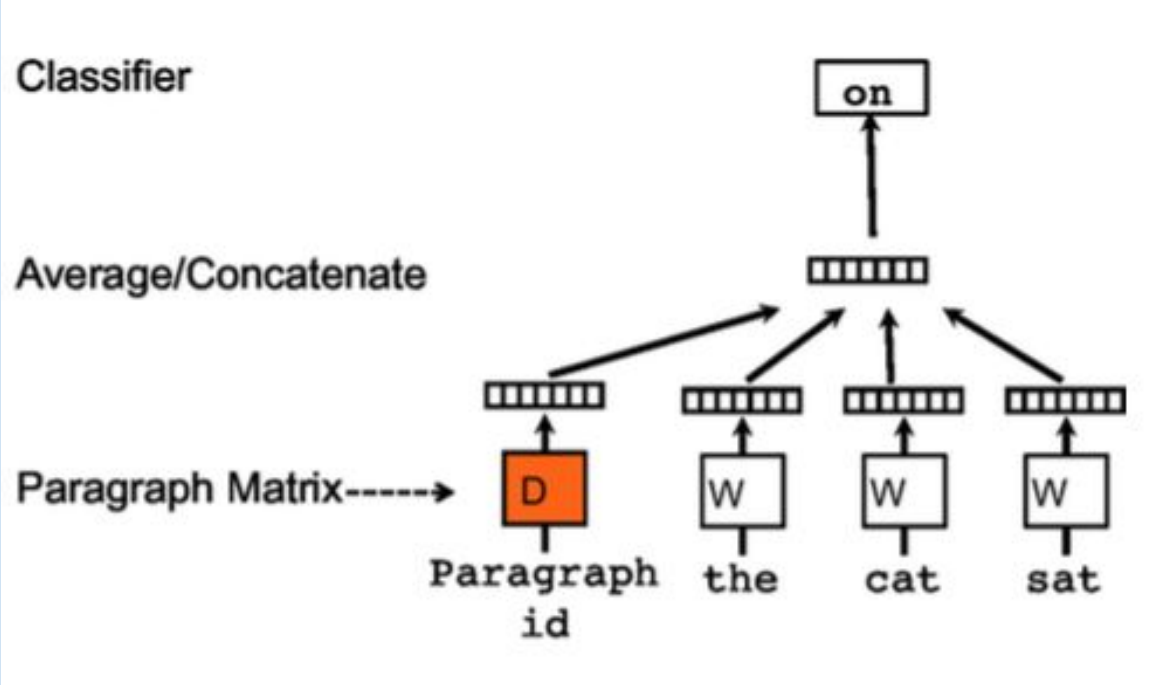
Drawbacks

1. High sensitivity to document length.
    - Long document sums wash out stronger signal points in FN classification.
2. “Nativity” Assumption in raw Sum:
    - Summation is commutative
    - Cannot capture word order or negation context:
        i. “Jill broke her crown but did not fall down.”
        ii. “Jill fell down but did not break her crown.”



P2V (Non-Naive Doc2Vec)

- P2V embed the full “paragraph” (doc) vector.
- Provides context for the whole paragraph.
- Tend to capture the semantics of the entire paragraph (doc).
- Some sensitivity to signal from word order.



Recurrent Neural Networks (RNNs)

RNNs look at data points **in order**

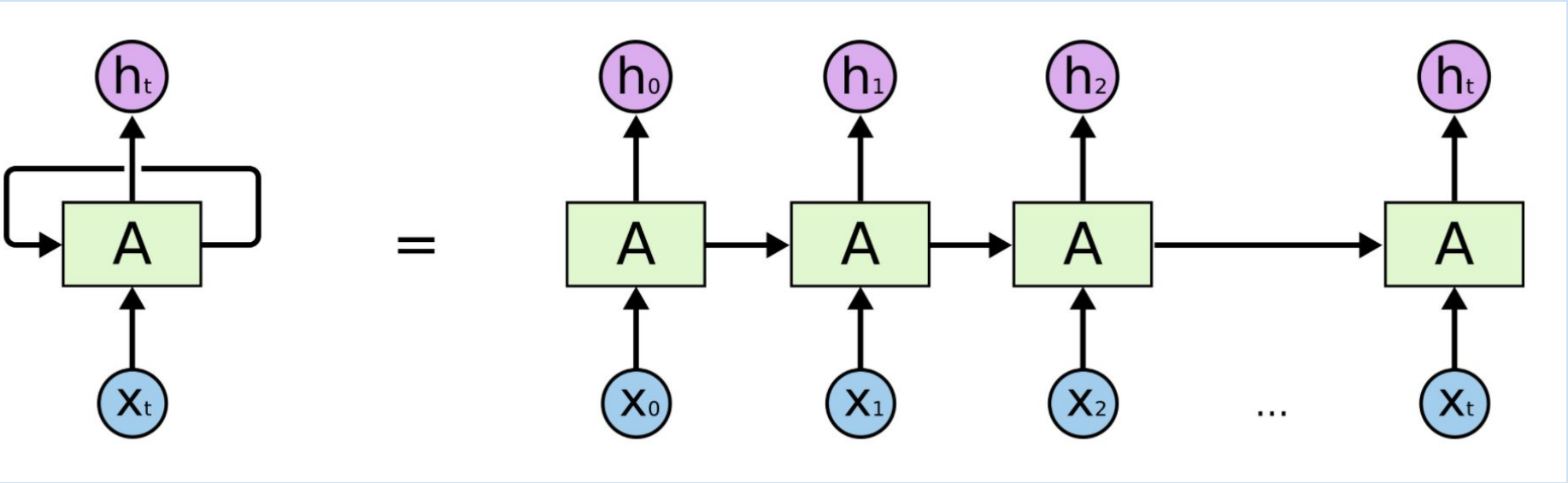
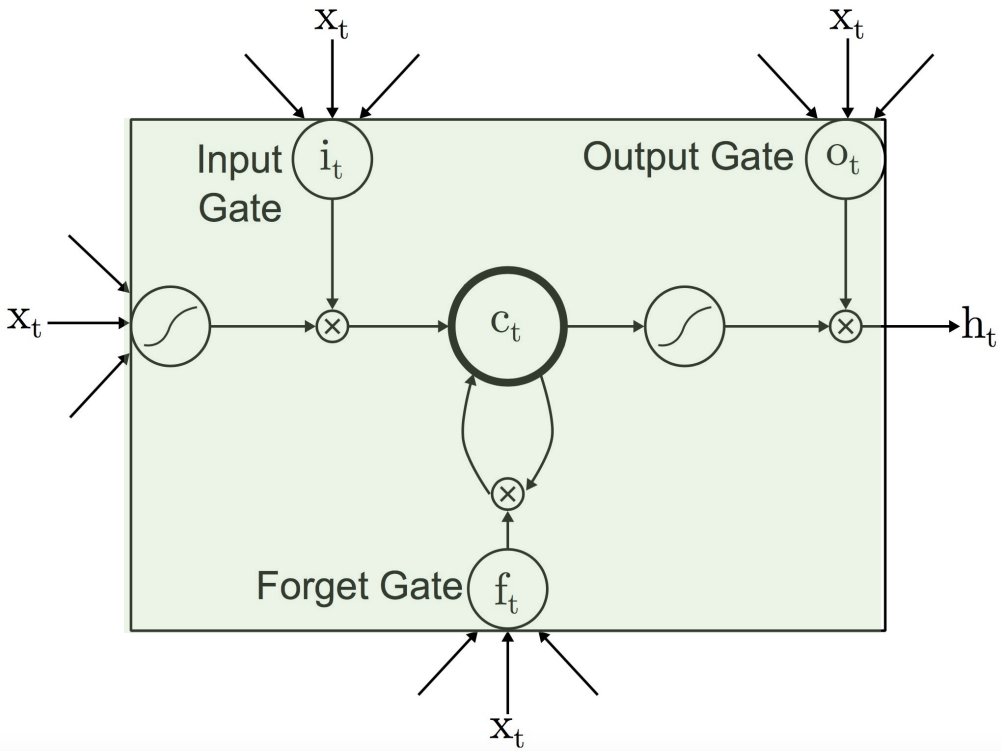


Image by Christopher Ohla

Long Short Term Memory (LSTM)

- `LSTM` RNNs are designed to capture context and track what it should "**remember**" and when it should "**forget**".



$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi} + b_i) \quad (1)$$

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf} + b_f) \quad (2)$$

$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho} + b_o) \quad (3)$$

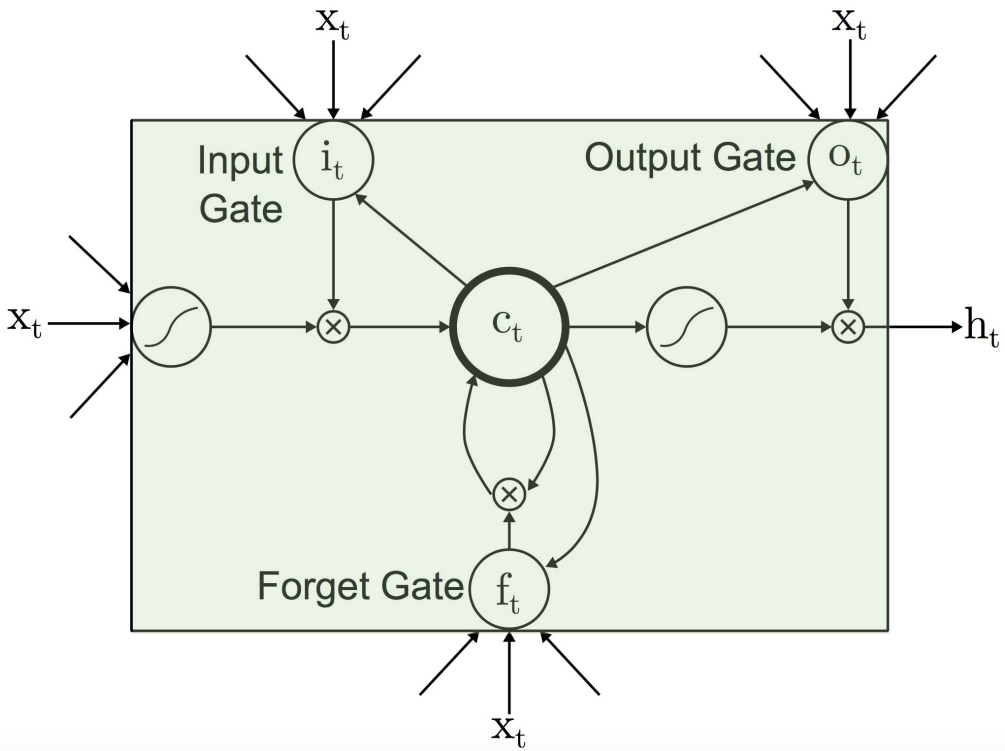
$$c_t = f_t \cdot c_{t-1} + i_t \sigma(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \quad (4)$$

$$h_t = o_t \cdot tanh(c_t) \quad (5)$$

*Neil et al (2016)

LSTM with Peep-holes

- Many variants of LSTMs offer high flexibility in model development.



$$i_t = \sigma(x_t W_{xi} + h_{t-1} W_{hi} + b_i + w_{ci} c_{t-1}) \quad (1)$$

$$f_t = \sigma(x_t W_{xf} + h_{t-1} W_{hf} + b_f + w_{cf} c_{t-1}) \quad (2)$$

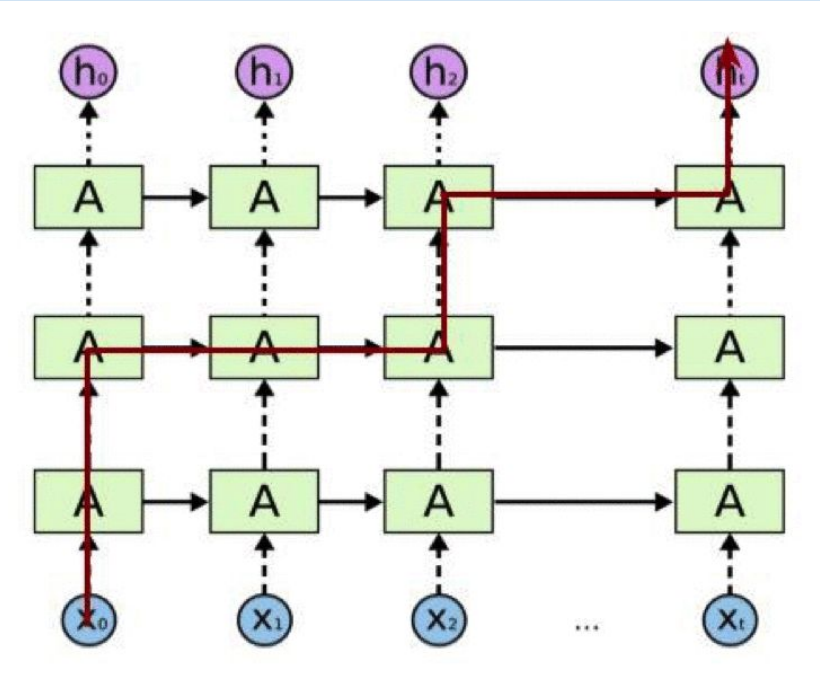
$$o_t = \sigma(x_t W_{xo} + h_{t-1} W_{ho} + b_o + w_{co} c_{t-1}) \quad (3)$$

$$c_t = f_t \cdot c_{t-1} + i_t \sigma(x_t W_{xc} + h_{t-1} W_{hc} + b_c) \quad (4)$$

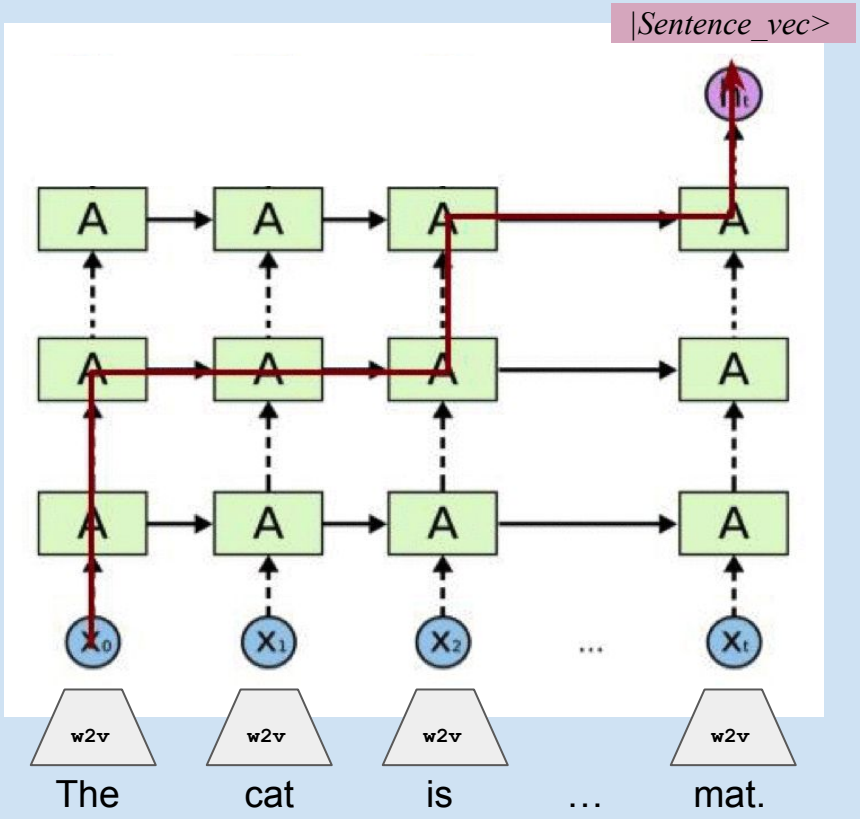
$$h_t = o_t \cdot tanh(c_t) \quad (5)$$

*Neil et al (2016)

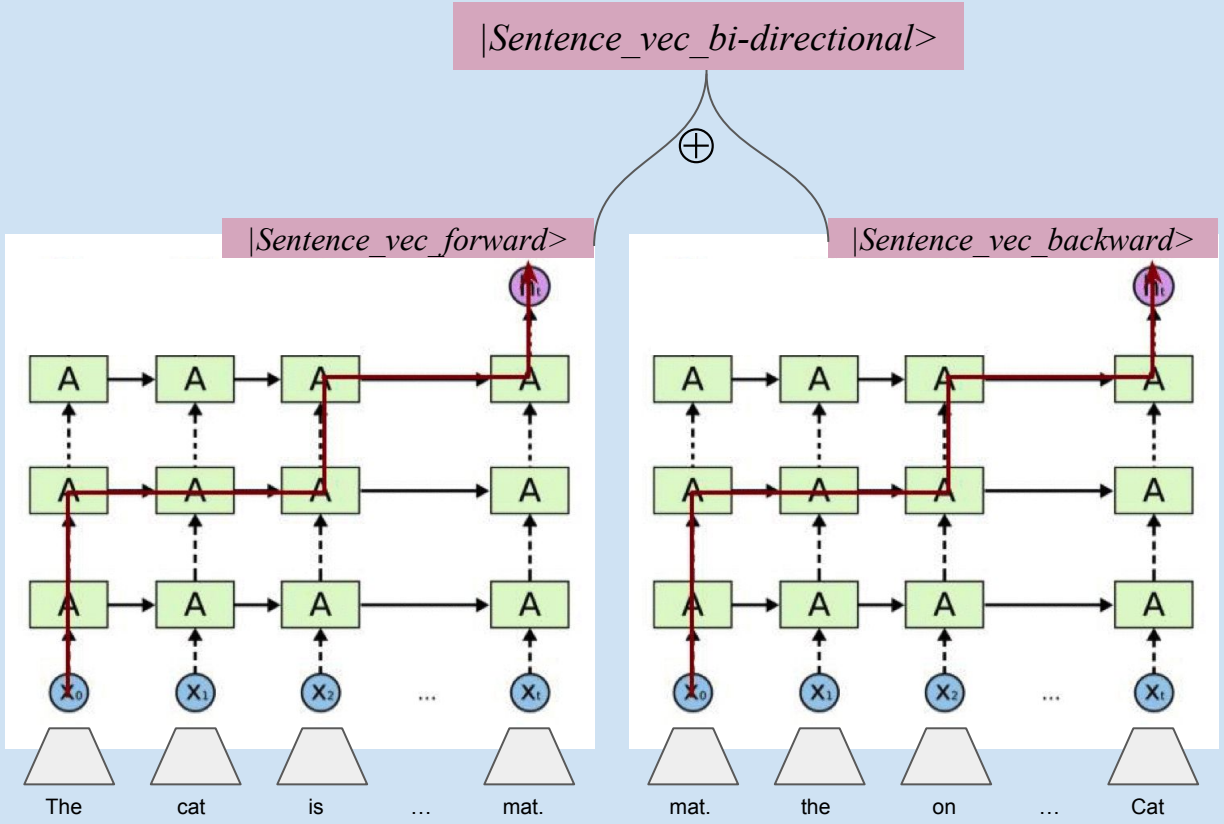
LSTM Stacking



LSTM Sentence-Vec Encoding



Bi-LSTM Encoding

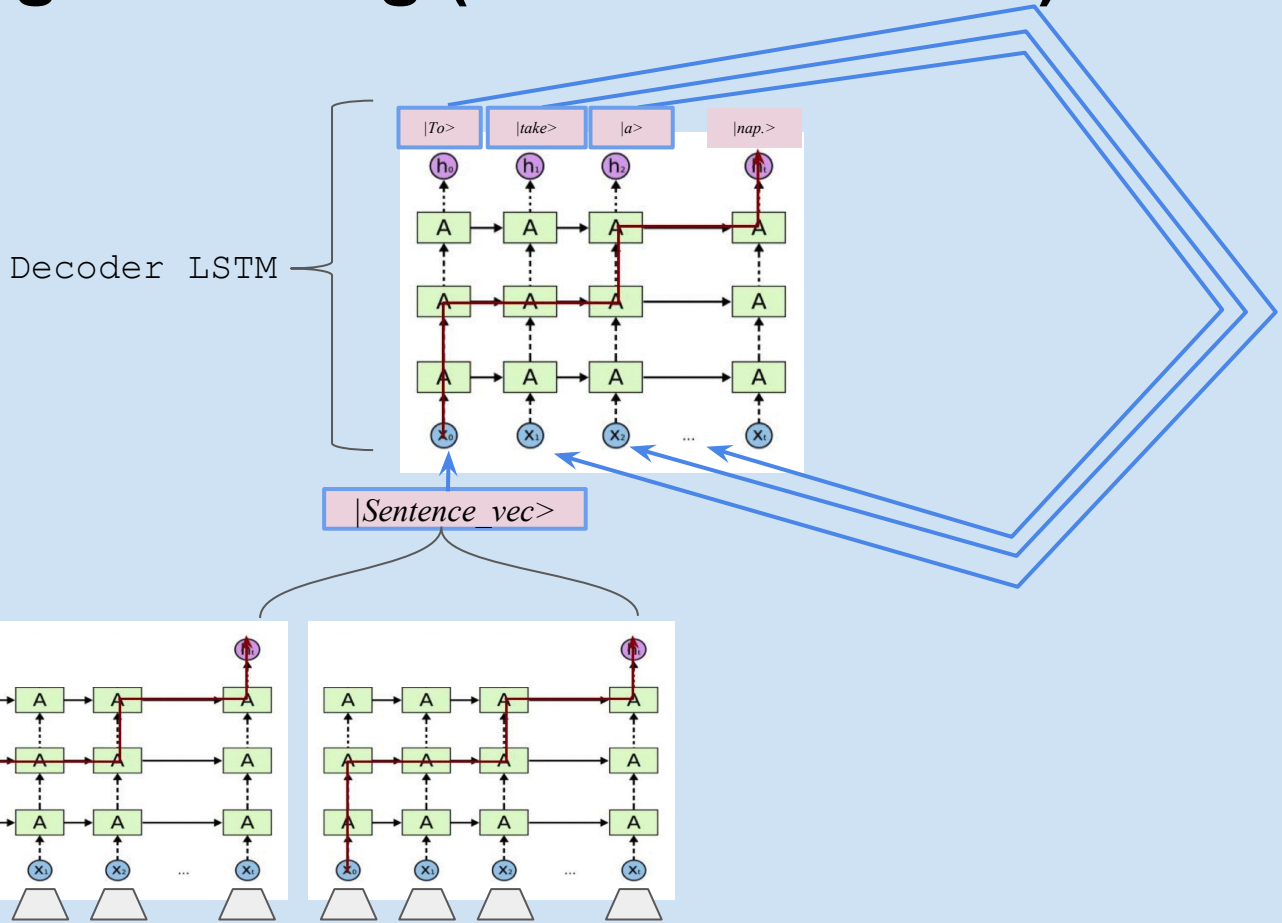


seq2seq Encoding-Decoding (Text Generation)

Bi-LSTMs encode semantics of the entire sentence.

Strong performance in several Use cases including...

- **Text generation**
- Translation
- Summarization...

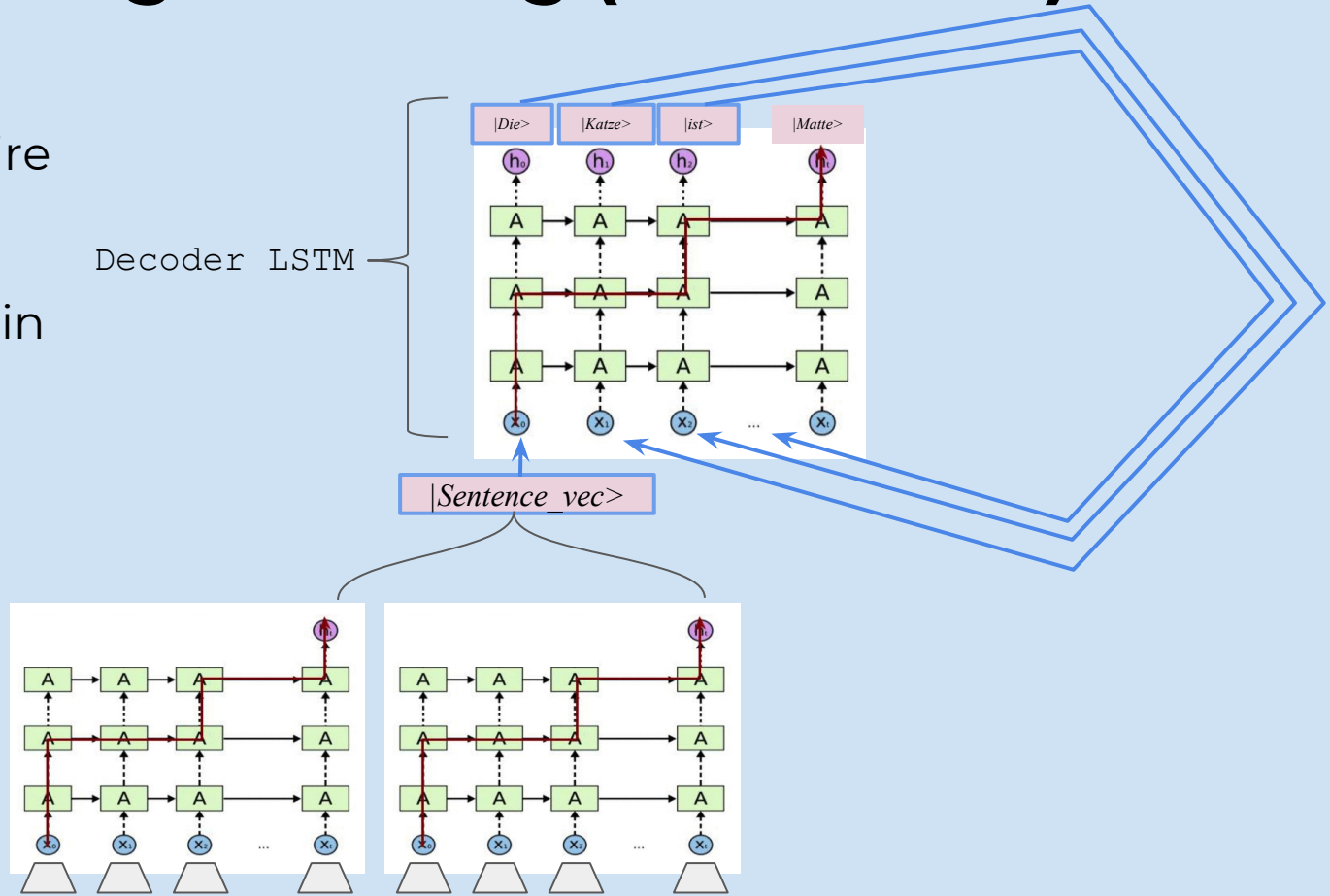


seq2seq Encoding-Decoding (Translation)

Bi-LSTMs encode semantics of the entire sentence.

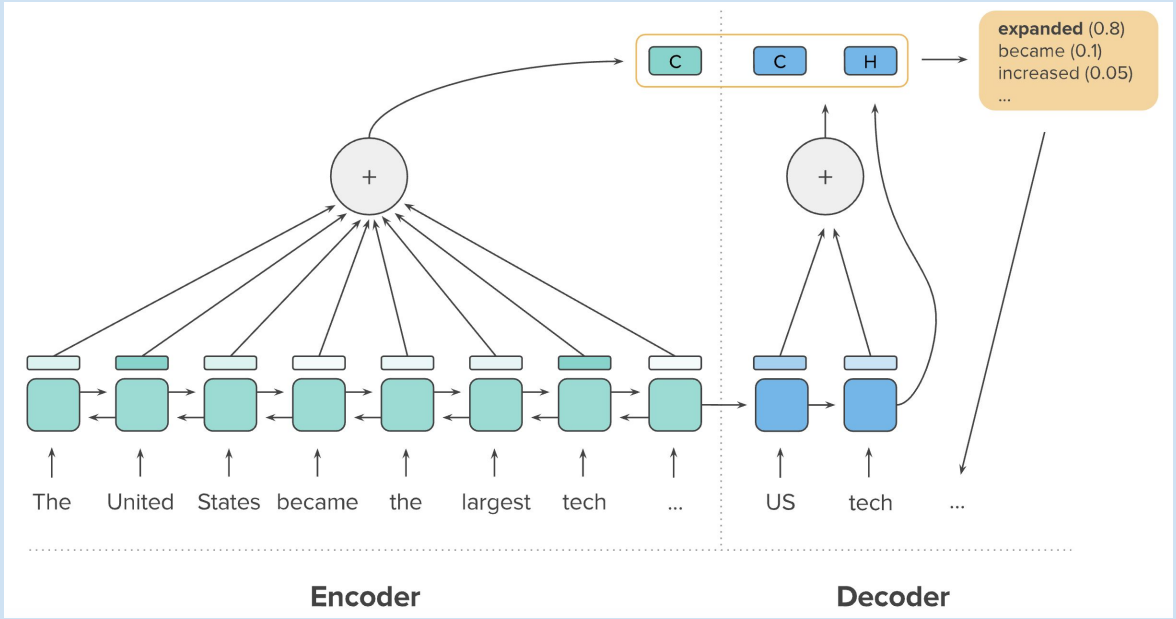
Strong performance in several Use cases including...

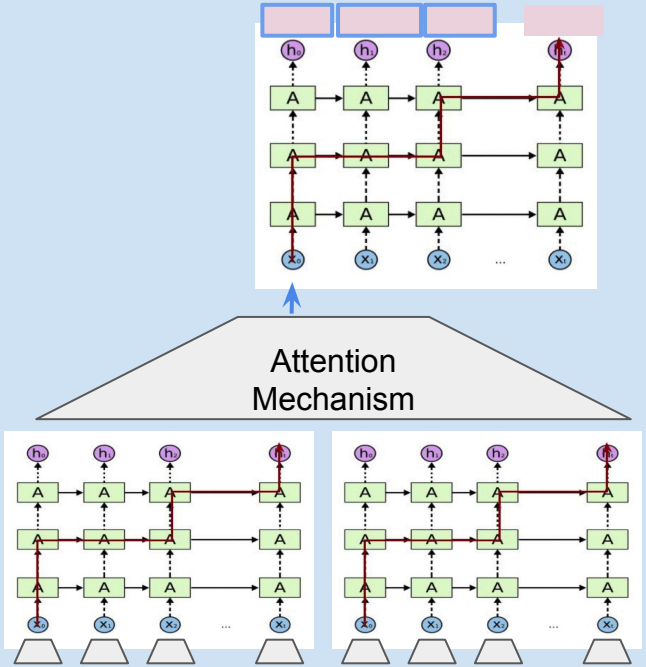
- Text generation
- **Translation**
- Summarization...



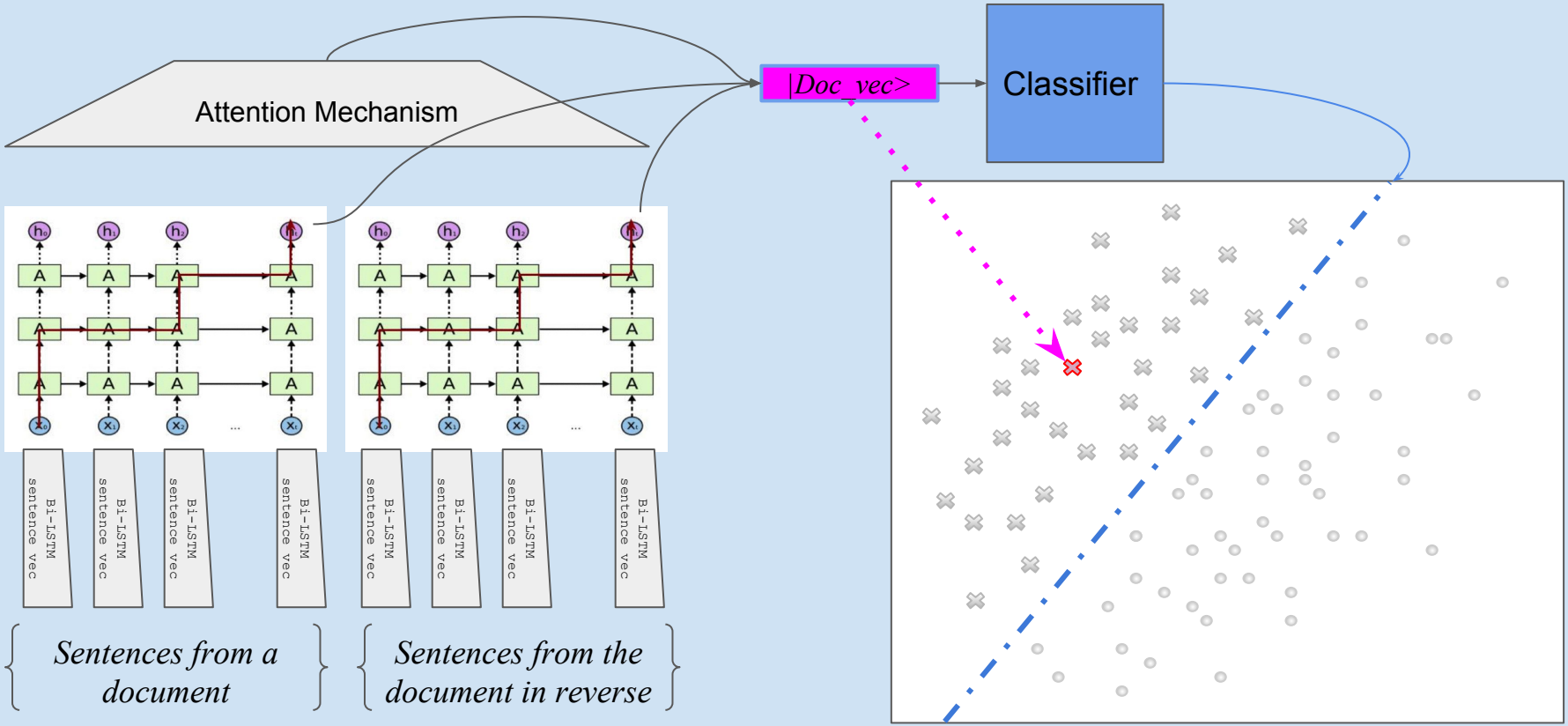
seq2seq Attention (Summarization)

- Attention mechanisms look at the outputs from each step of the `Bi-LSTM` network and learn to leverage the entire text sequence.





Bi-LSTM doc2vec Classification



Solving the Problem: Applying the Algorithm

(Remember) Solving “Fake News” means…

Image Source

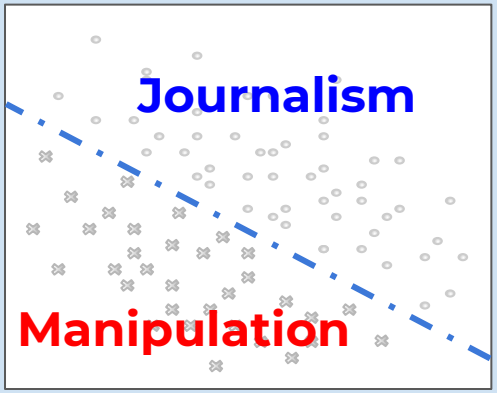
Enable readers to detect when the article’s goal is

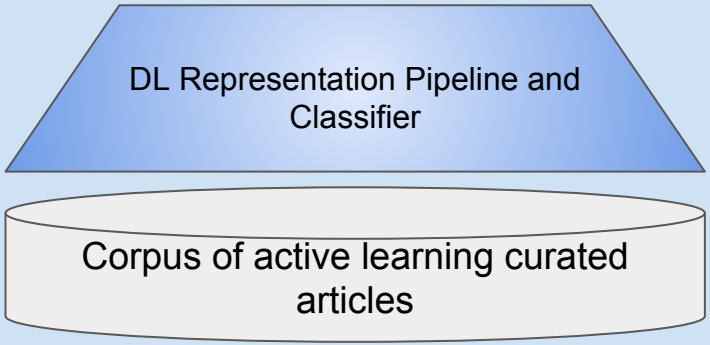
credible information sharing

vs.

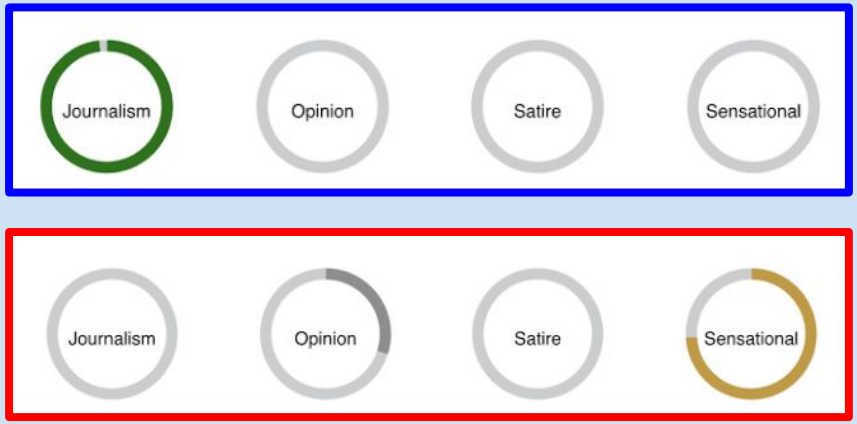
emotional manipulation.

Applying the Algorithm



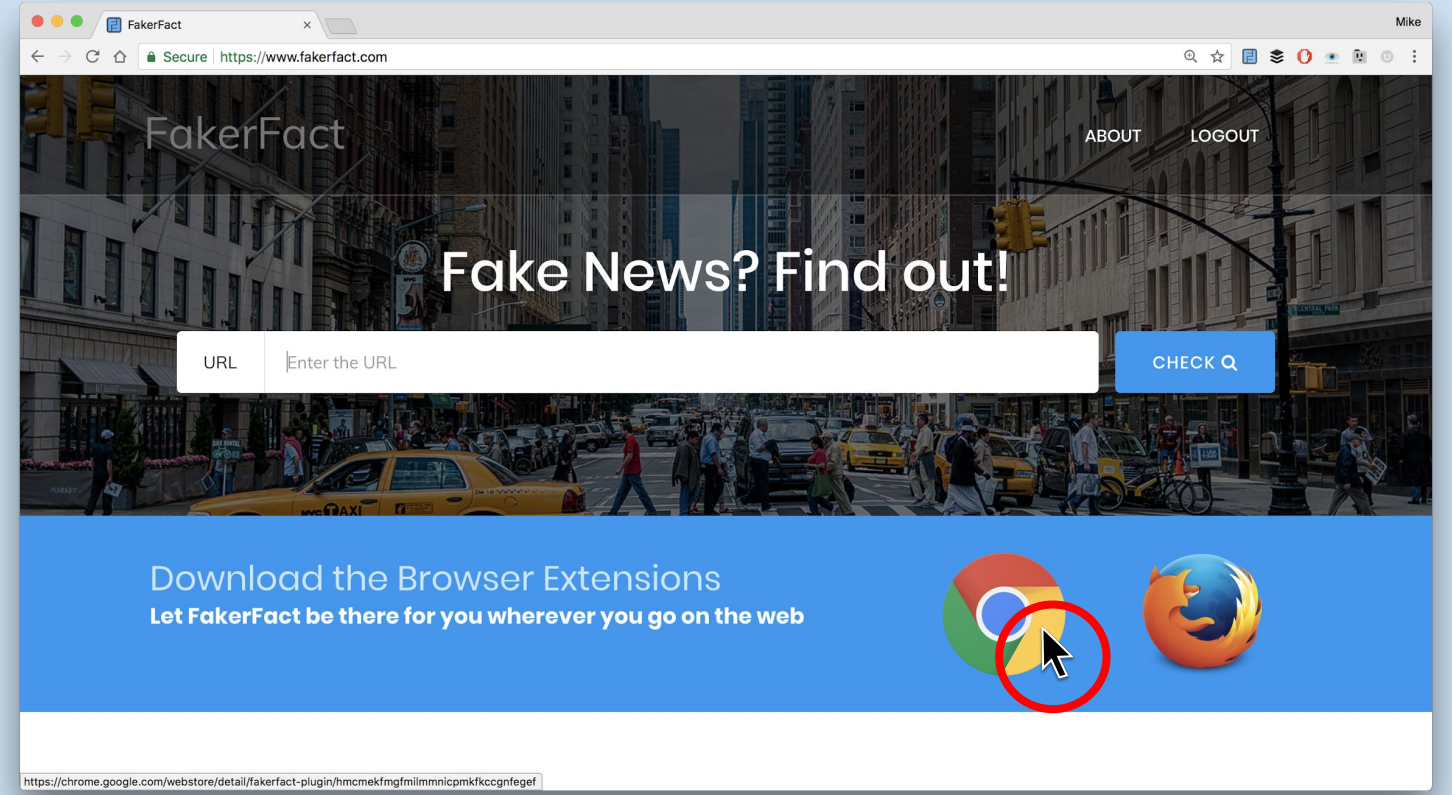


1. **Training:** $O(10^6)$ of active learning curated articles for clear signal.
2. Use DL techniques to generate rich vector representation of articles.
3. Classification:

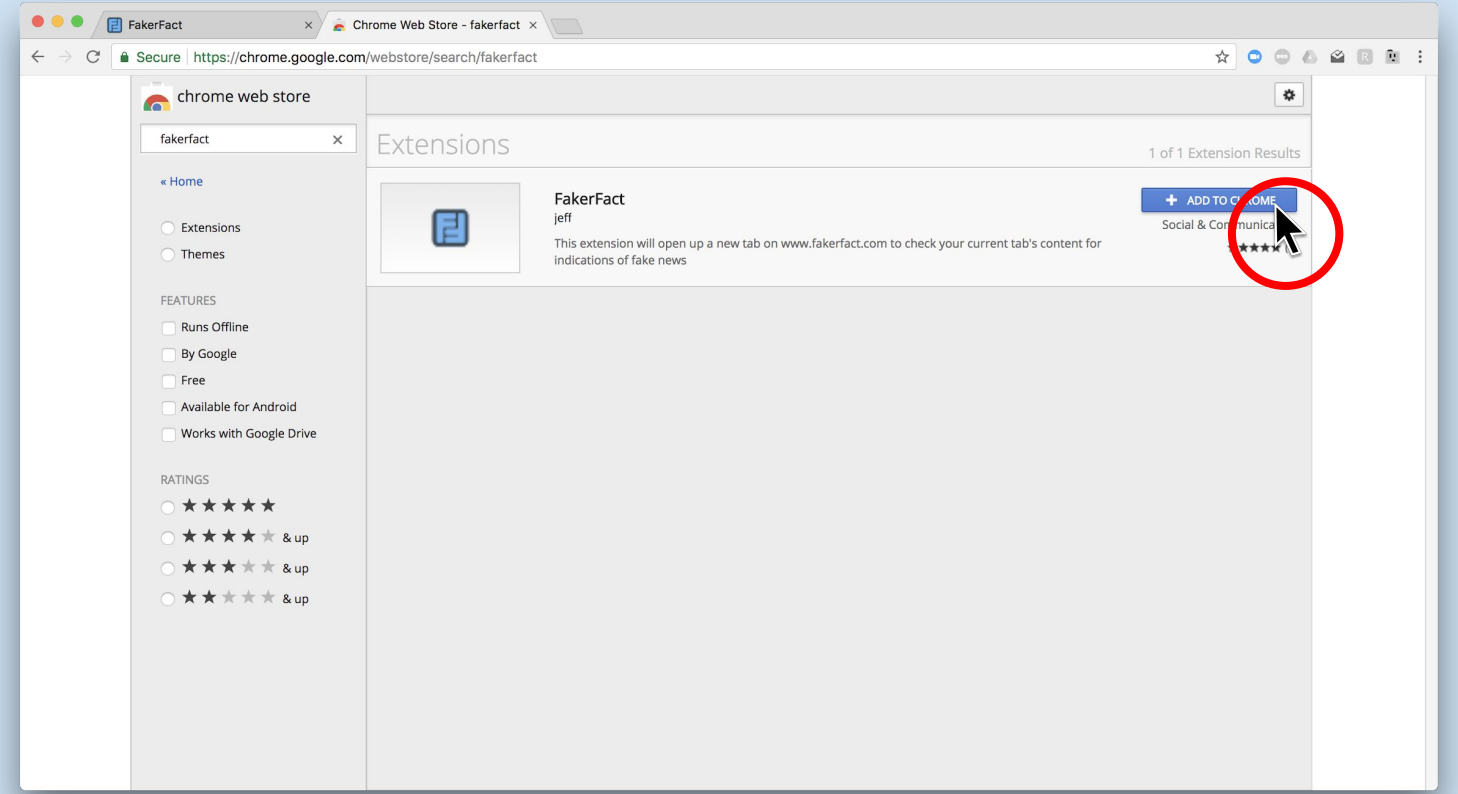


Algorithm in Action

FakerFact


FakerFact
Mike
Secure https://www.fakerfact.com
FakerFact
ABOUT
LOGOUT
Fake News? Find out!
URL
Enter the URL
CHECK
Download the Browser Extensions
Let FakerFact be there for you wherever you go on the web
https://chrome.google.com/webstore/detail/fakerfact-plugin/hmcmekfmgfmilmmnicpmkfkccgnfegef

FakerFact


FakerFact
Chrome Web Store - fakerfact
Secure
https://chrome.google.com/webstore/search/fakerfact
chrome web store
fakerfact
« Home
Extensions
Themes
FEATURES
Runs Offline
By Google
Free
Available for Android
Works with Google Drive
RATINGS
& up
& up
& up
Extensions
1 of 1 Extension Results
FakerFact
jeff
This extension will open up a new tab on www.fakerfact.com to check your current tab's content for indications of fake news
ADD TO CHROME
Social & Communica

FakerFact

Trump Readies Sweeping Tariff

Mike

Secure | https://www.nytimes.com/2018/03/15/us/politics/trump-china-trade-measures.html?hp&action=click&pgtype=Homepage&clickSource=story-heading&...

POLITICS | Trump Readies Sweeping Tariffs and Investment Restrictions on China

432

Trump Readies Sweeping Tariffs and Investment Restrictions on China

By ANA SWANSON MARCH 15, 2018

A factory for the smartphone maker Oppo in Dongguan, China. The tariffs could extend to more mundane products, including consumer electronics, apparel and even shoes. Nicolas Asfouri/Getty Images

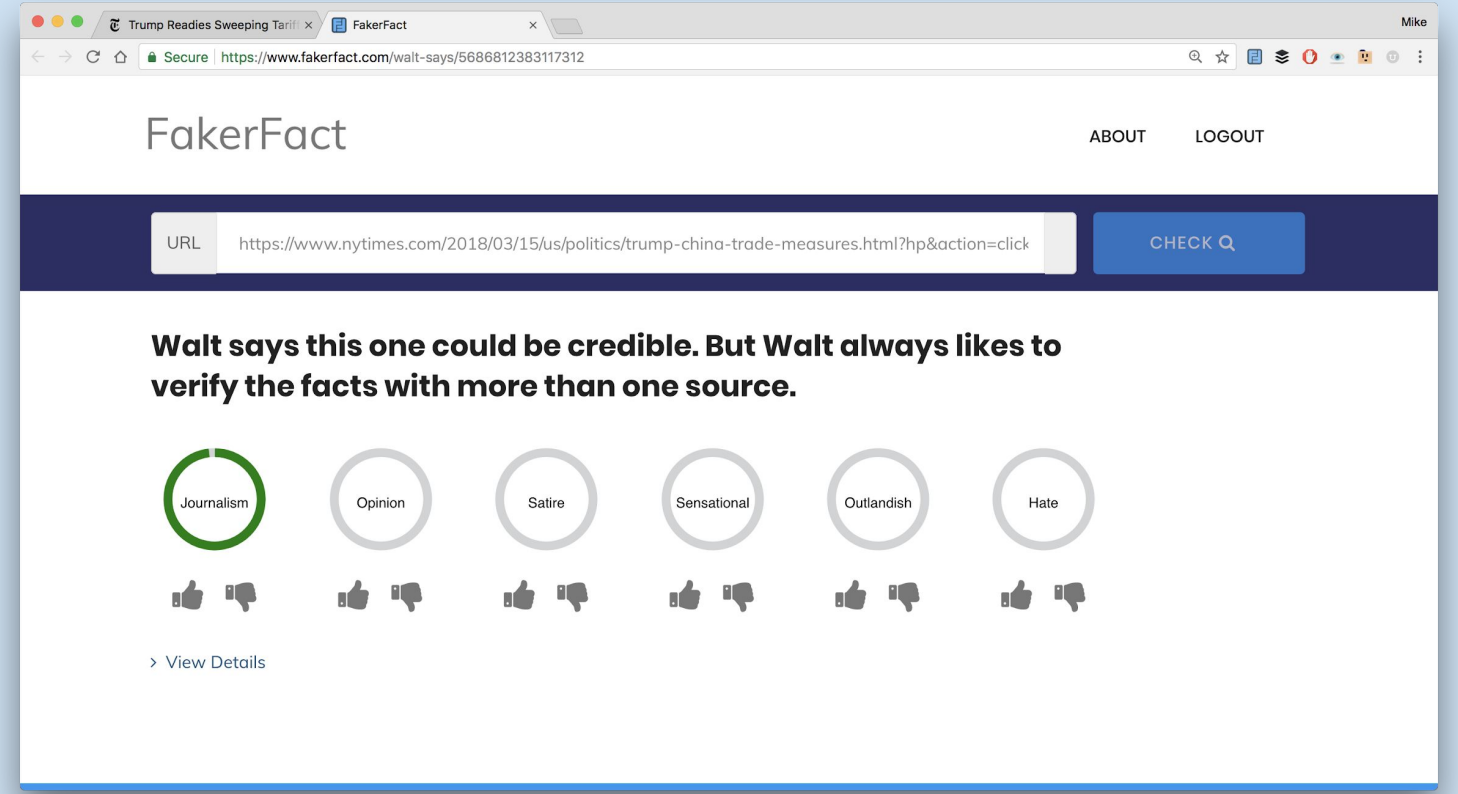
The Trump White House

The historic moments, head-spinning developments and inside-the-White House intrigue.

- Newly Emboldened, Trump Says What He Really Feels — MAR 18
- Before Saudi's Visit, Administration Implores Congress Not to Block Military Aid — MAR 18
- Army, Struggling to Get Technology in Soldiers' Hands, Tries the Unconventional — MAR 18
- Andrew McCabe, Fired F.B.I. Deputy, Is Said to Have Kept Memos on Trump — MAR 17
- Smuggling of U.S. Technology Is Outpacing Cold War Levels, Experts Say — MAR 17

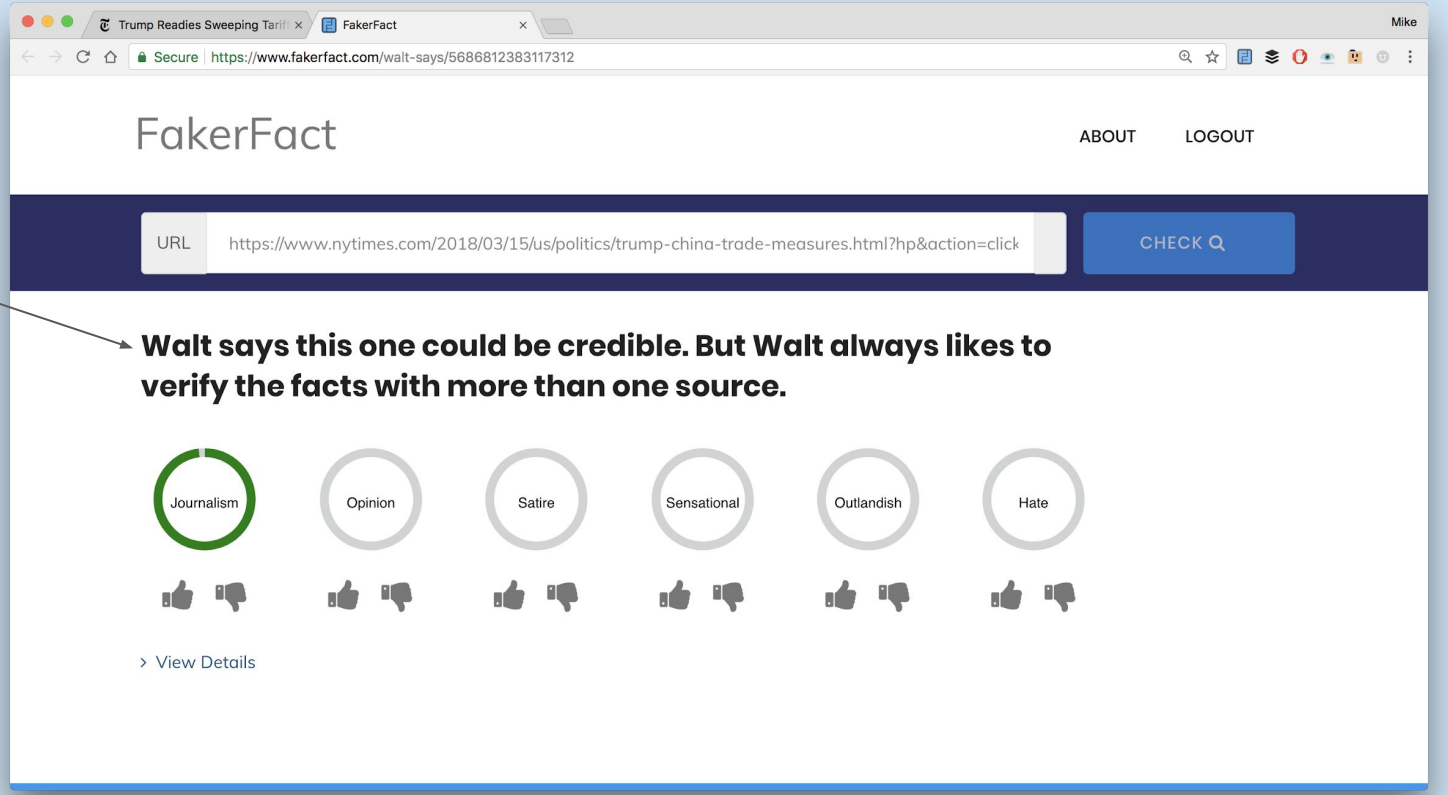
See More »

FakerFact


Trump Readies Sweeping Tariff
FakerFact
Mike
Secure
https://www.fakerfact.com/walt-says/5686812383117312
FakerFact
ABOUT
LOGOUT
URL
https://www.nytimes.com/2018/03/15/us/politics/trump-china-trade-measures.html?hp&action=click
CHECK
Walt says this one could be credible. But Walt always likes to verify the facts with more than one source.
Journalism
Opinion
Satire
Sensational
Outlandish
Hate
View Details

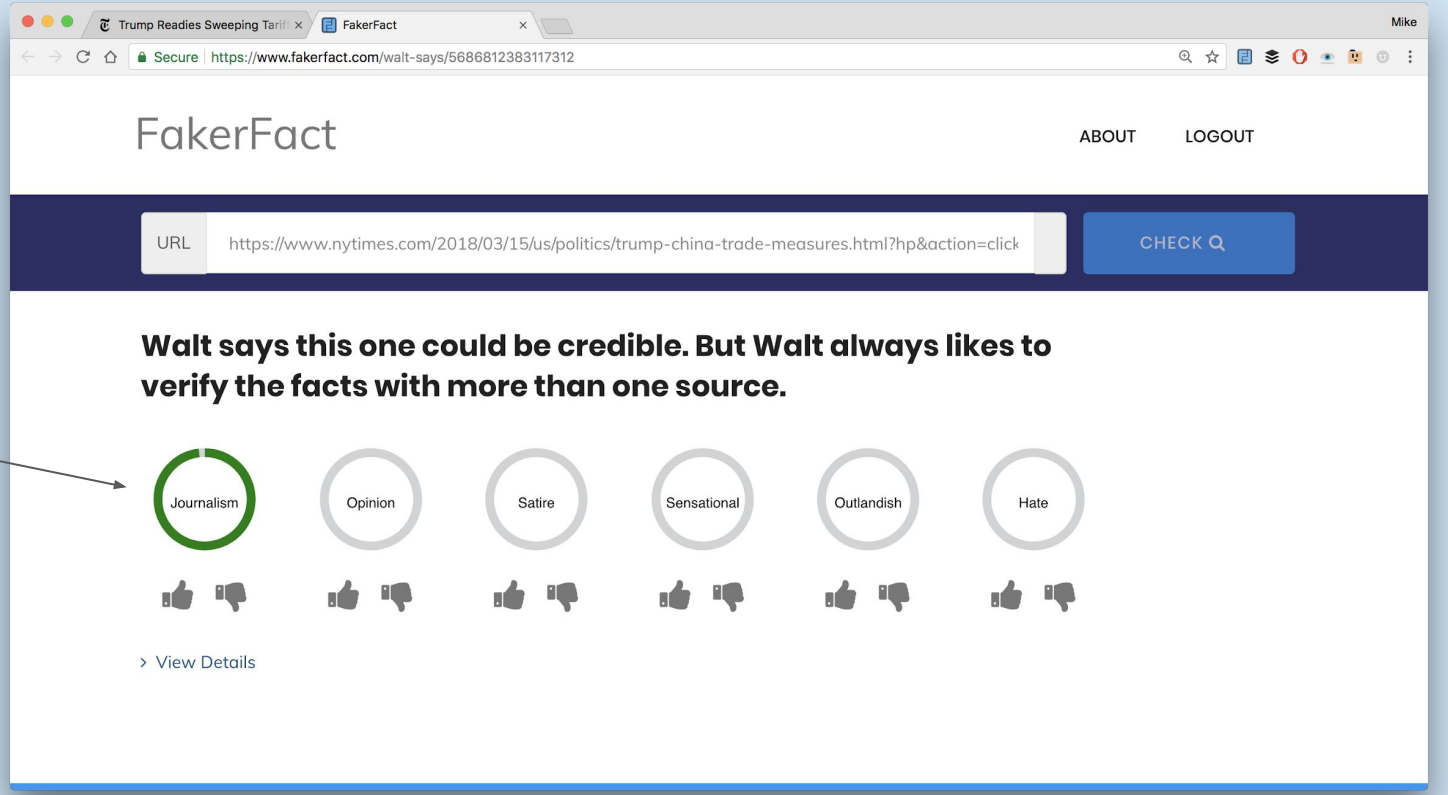
FakerFact

Every result has a “Walt says” summary of the scores.



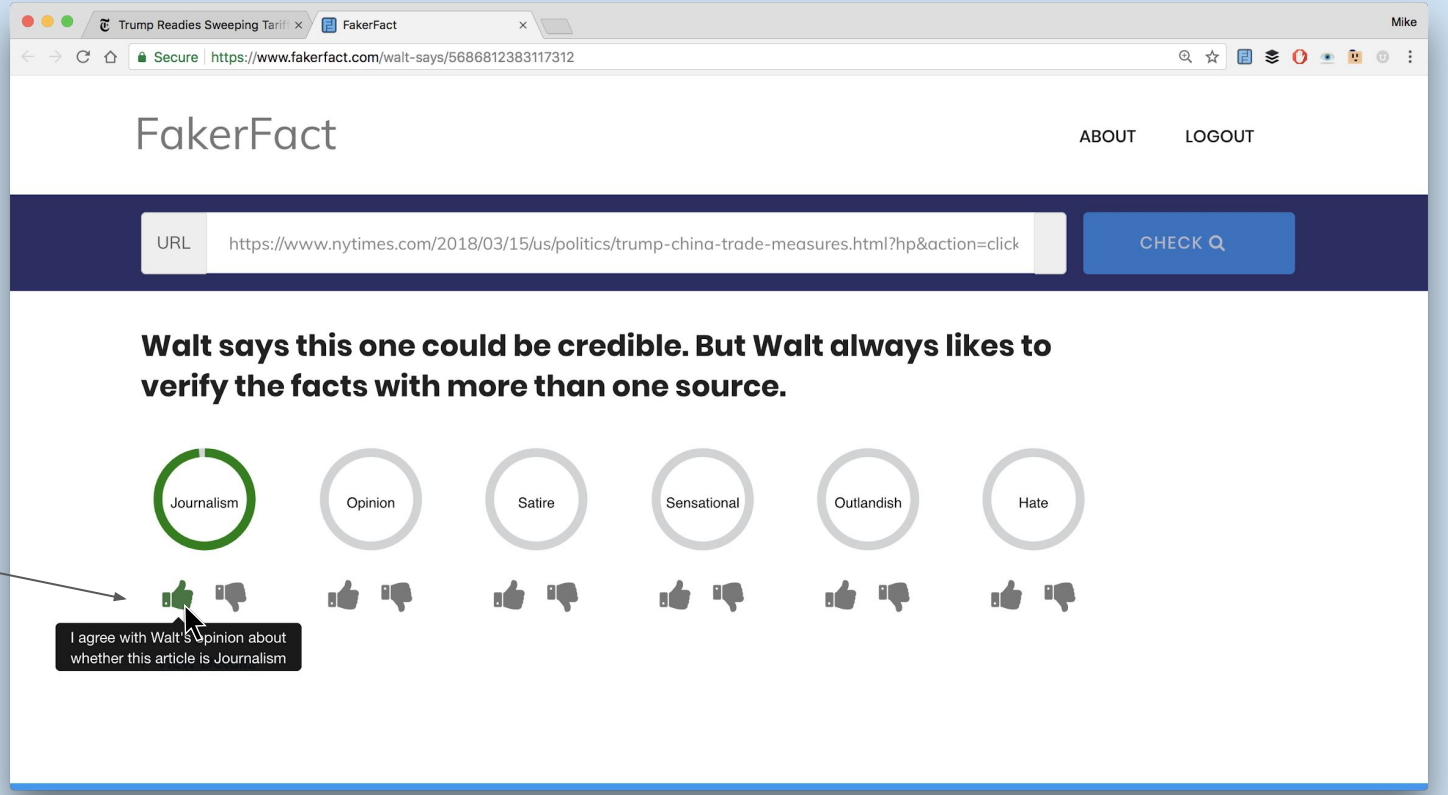
FakerFact

Detailed analysis of how the article scores in several dimensions of news credibility.

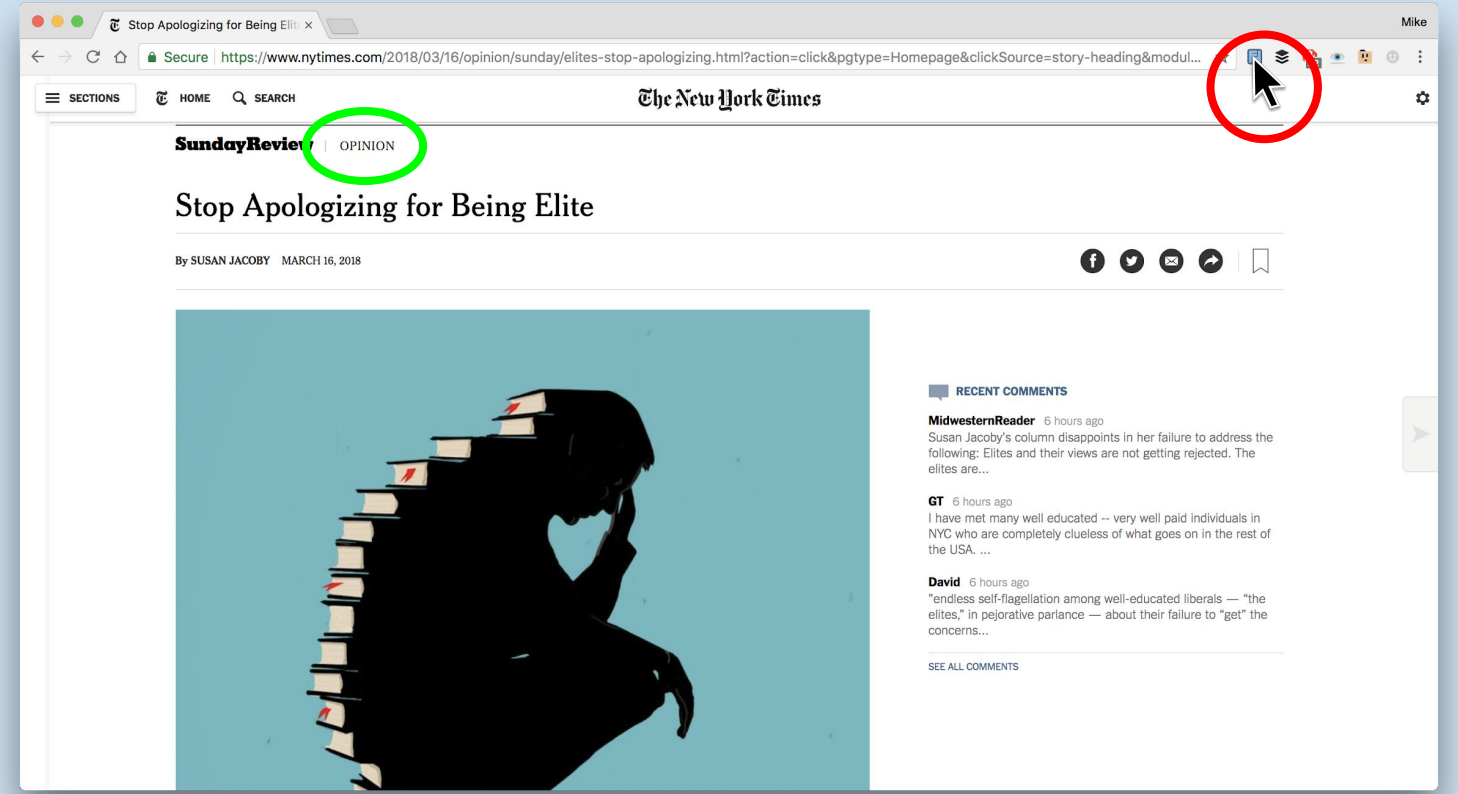


FakerFact

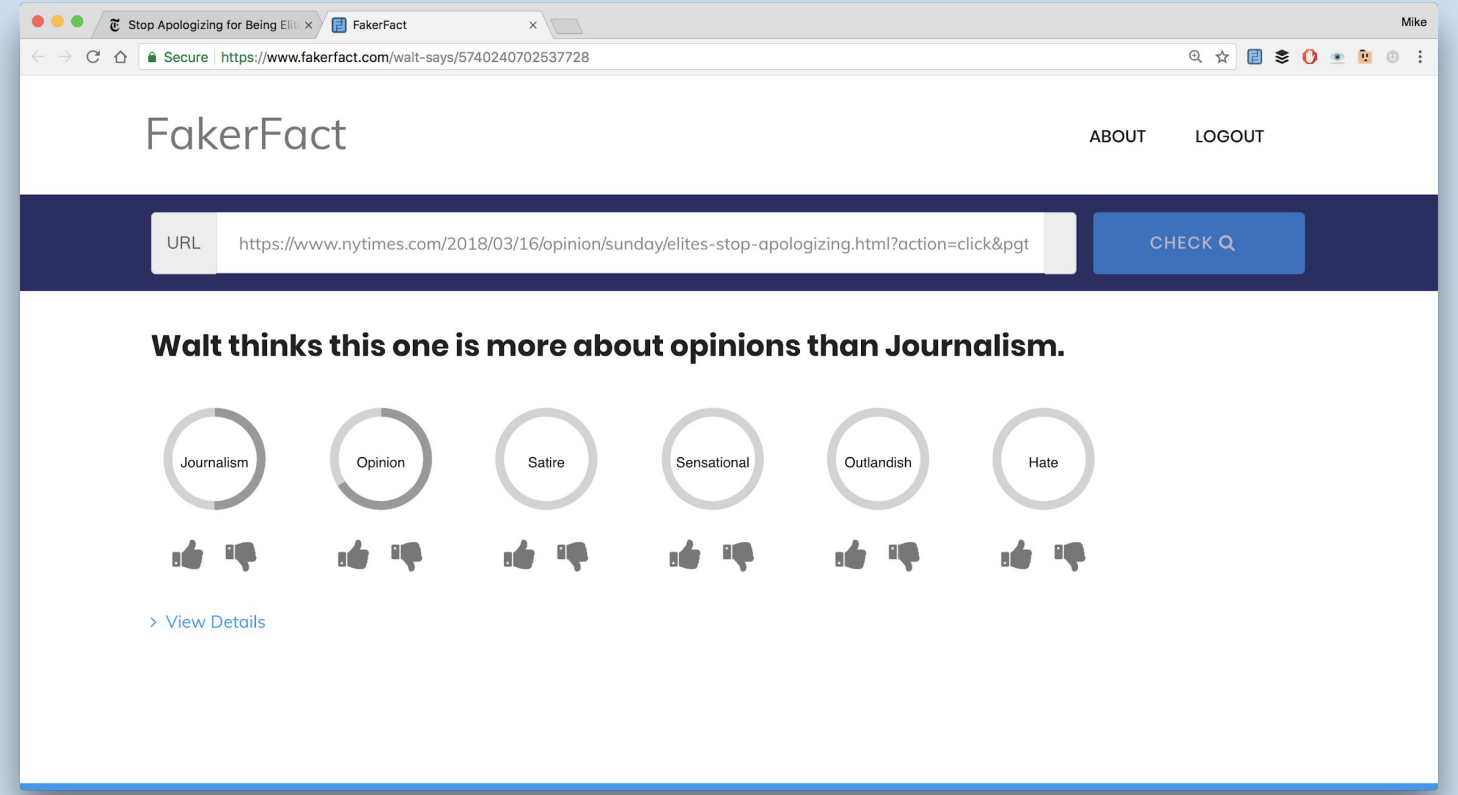
Opportunity for feedback on any results for model fine tuning.



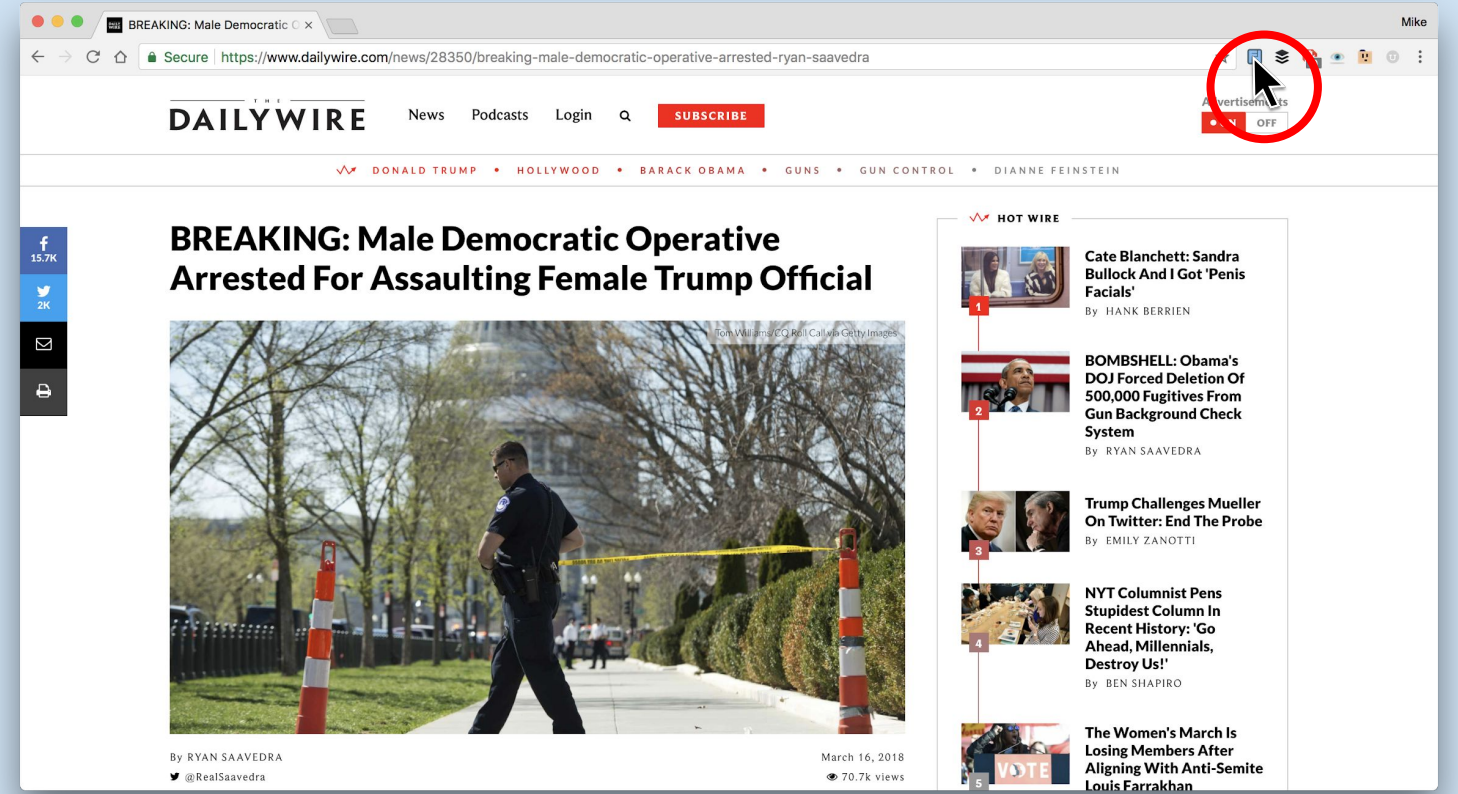
FakerFact


Stop Apologizing for Being Elit
Mike
Secure https://www.nytimes.com/2018/03/16/opinion/sunday/elites-stop-apologizing.html?action=click&pgtype=Homepage&clickSource=story-heading&modul...
SECTIONS
HOME
SEARCH
The New York Times
SundayReview | OPINION
Stop Apologizing for Being Elite
By SUSAN JACOBY MARCH 16, 2018
RECENT COMMENTS
MidwesternReader 6 hours ago
Susan Jacoby's column disappoints in her failure to address the following: Elites and their views are not getting rejected. The elites are...
GT 6 hours ago
I have met many well educated -- very well paid individuals in NYC who are completely clueless of what goes on in the rest of the USA. ...
David 6 hours ago
"endless self-flagellation among well-educated liberals — "the elites," in pejorative parlance — about their failure to "get" the concerns...
SEE ALL COMMENTS

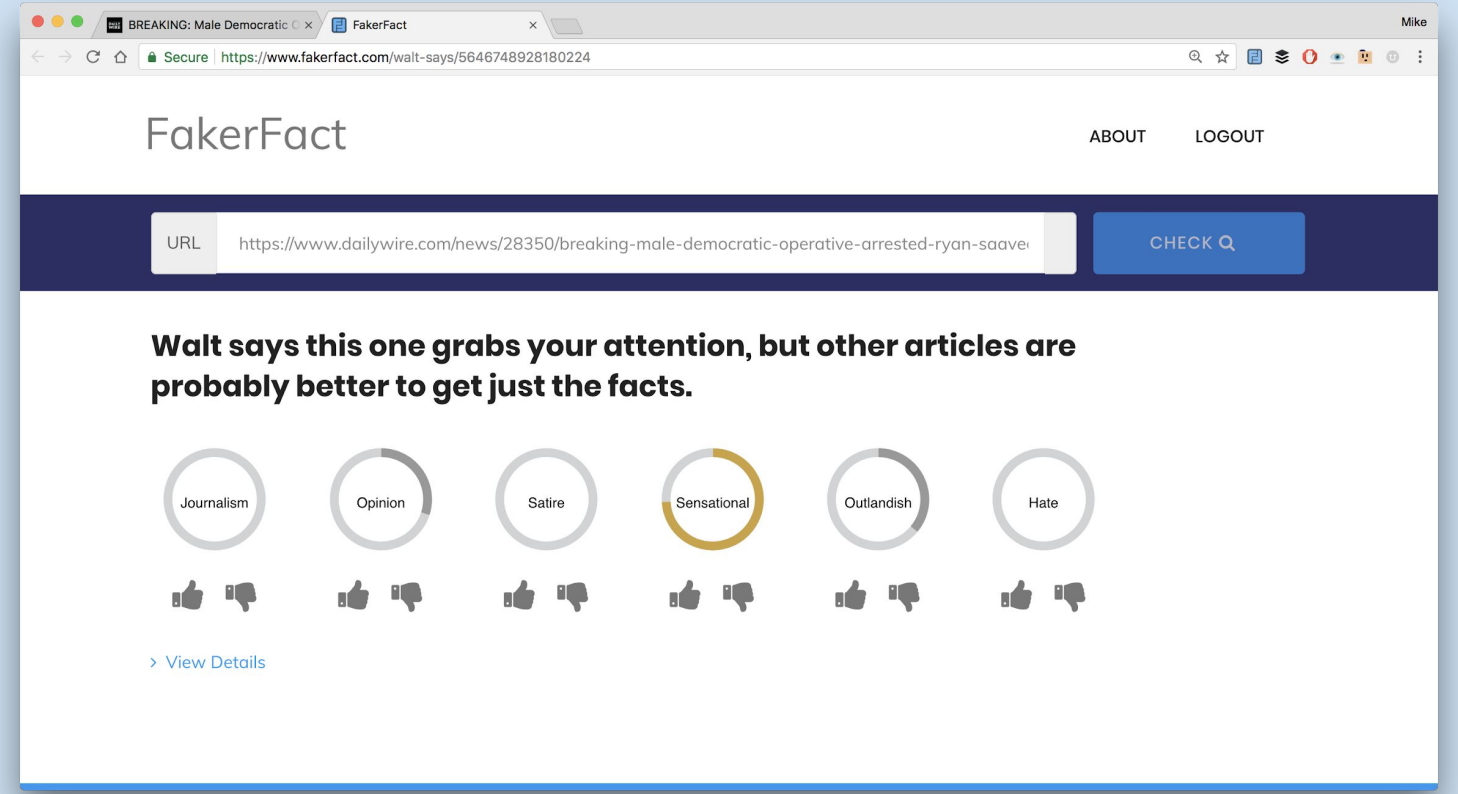
FakerFact


Stop Apologizing for Being Elit
FakerFact
Mike
Secure https://www.fakerfact.com/walt-says/5740240702537728
FakerFact
ABOUT
LOGOUT
URL
https://www.nytimes.com/2018/03/16/opinion/sunday/elites-stop-apologizing.html?action=click&pgt
CHECK
Walt thinks this one is more about opinions than Journalism.
Journalism
Opinion
Satire
Sensational
Outlandish
Hate
View Details

FakerFact


BREAKING: Male Democratic C
Mike
Secure https://www.dailywire.com/news/28350/breaking-male-democratic-operative-arrested-ryan-saavedra
DAILYWIRE
News
Podcasts
Login
SUBSCRIBE
Advertisements
ON
OFF
DONALD TRUMP • HOLLYWOOD • BARACK OBAMA • GUNS • GUN CONTROL • DIANNE FEINSTEIN
15.7K
2K
BREAKING: Male Democratic Operative Arrested For Assaulting Female Trump Official
Tom Williams/CQ Roll Call via Getty Images
By RYAN SAAVEDRA
@RealSaavedra
March 16, 2018
70.7k views
HOT WIRE
1
Cate Blanchett: Sandra Bullock And I Got 'Penis Facials'
By HANK BERRIEN
2
BOMBSHELL: Obama's DOJ Forced Deletion Of 500,000 Fugitives From Gun Background Check System
By RYAN SAAVEDRA
3
Trump Challenges Mueller On Twitter: End The Probe
By EMILY ZANOTTI
4
NYT Columnist Pens Stupidest Column In Recent History: 'Go Ahead, Millennials, Destroy Us!'
By BEN SHAPIRO
5
The Women's March Is Losing Members After Aligning With Anti-Semite Louis Farrakhan

FakerFact


BREAKING: Male Democratic C
FakerFact
Mike
Secure https://www.fakerfact.com/walt-says/5646748928180224
FakerFact
ABOUT
LOGOUT
URL
https://www.dailywire.com/news/28350/breaking-male-democratic-operative-arrested-ryan-saave
CHECK
Walt says this one grabs your attention, but other articles are probably better to get just the facts.
Journalism
Opinion
Satire
Sensational
Outlandish
Hate
View Details