Fake News Detection

Team 39
Iris Yoon, Postdoc Applied Mathematics, University of Pennsylvania
Rabiya Noori, MSc Biomedical Engineering, University of Toronto
Jerri Zhang, MS Health Data Science, Harvard University
Renee G. Reynolds, Master of Financial Engineering, UC Berkeley
Hannah Mei, BA Mathematics and Economics, NYU

October 16, 2020

Misinformation and Media

➔ One-in-five US adults get their political news primarily through social media.

Impact of Fake News

➔ Social: Fake news spread faster and further than the truth

➔ Economic: $130 billion loss in market cap after hacked AP tweet

➔ Business: Customer trust in social media platforms

An Infeasible Solution

➔ Hiring professionals for fact checking is costly

Solution

Develop a model to classify news articles as Real or Fake

Agenda

1. **Exploratory Data Analysis**
2. **Models & Training**
    a. Baseline Model - Recurrent Neural Networks (RNN)
    b. Advanced Model - Bidirectional Encoder Representations from Transformers (BERT)
3. **Model Interpretability**
    a. Local Interpretable Model-Agnostic Explanations (LIME)
4. **Conclusion & Future Work**

Data Wrangling & Cleaning

Data: 20,000 fake and 20,000 real news articles (kaggle)

Example

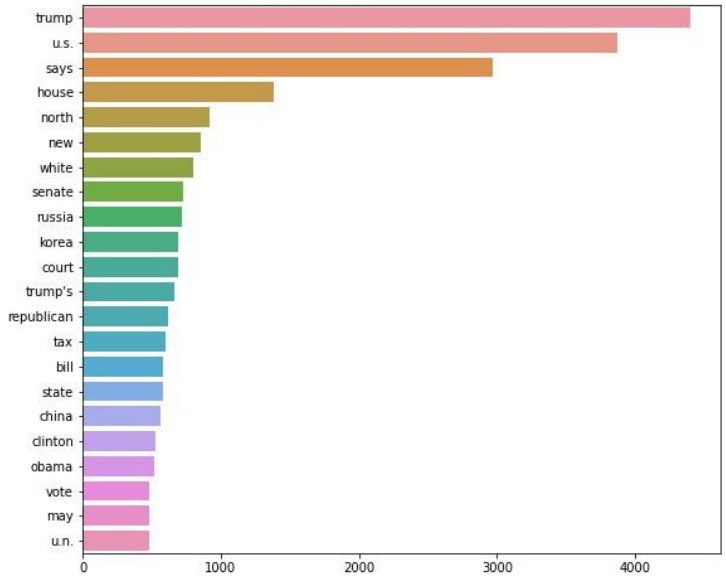
Fake	Real
Title: Donald Trump Just Visited The 9/11 Museum And P*ssed EVERYONE Off(IMAGES/TWEETS) **Text:** Native New Yorker and Republican front runner Donald Trump made his first-ever visit ... Trump doesn t give a damn about the victims, survivors or families affected by 9/11 he only cares about making himself look better. Featured image via Josh Andew Instagram	**Title:** White House, Congress prepare for talks on spending, immigration **Text:** WASHINGTON (Reuters) - The White House said on Friday it was set to kick off talks ... Trump, who won his first major legislative victory with the passage of a major tax overhaul this month, has also promised a major infrastructure plan.

Data cleaning

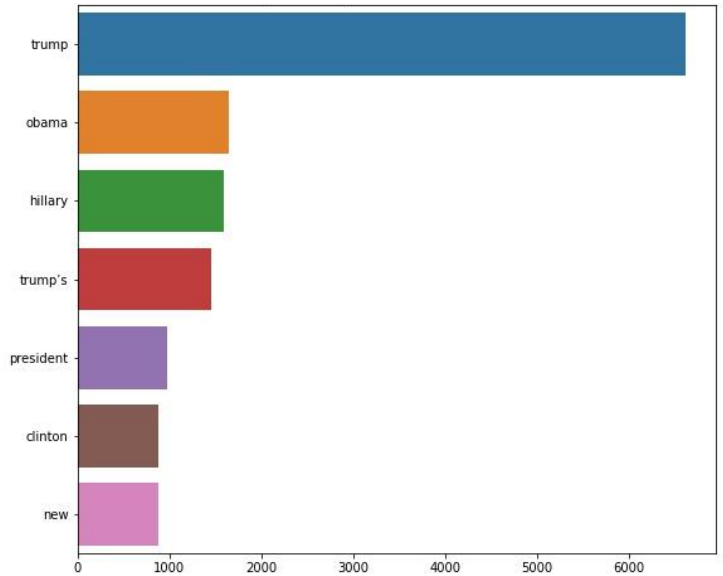
Exploratory Data Analysis

Most popular words in news article headlines (excluding stopwords, e.g. ‘is’, ‘of’, ‘be’, ‘to’)

Real News



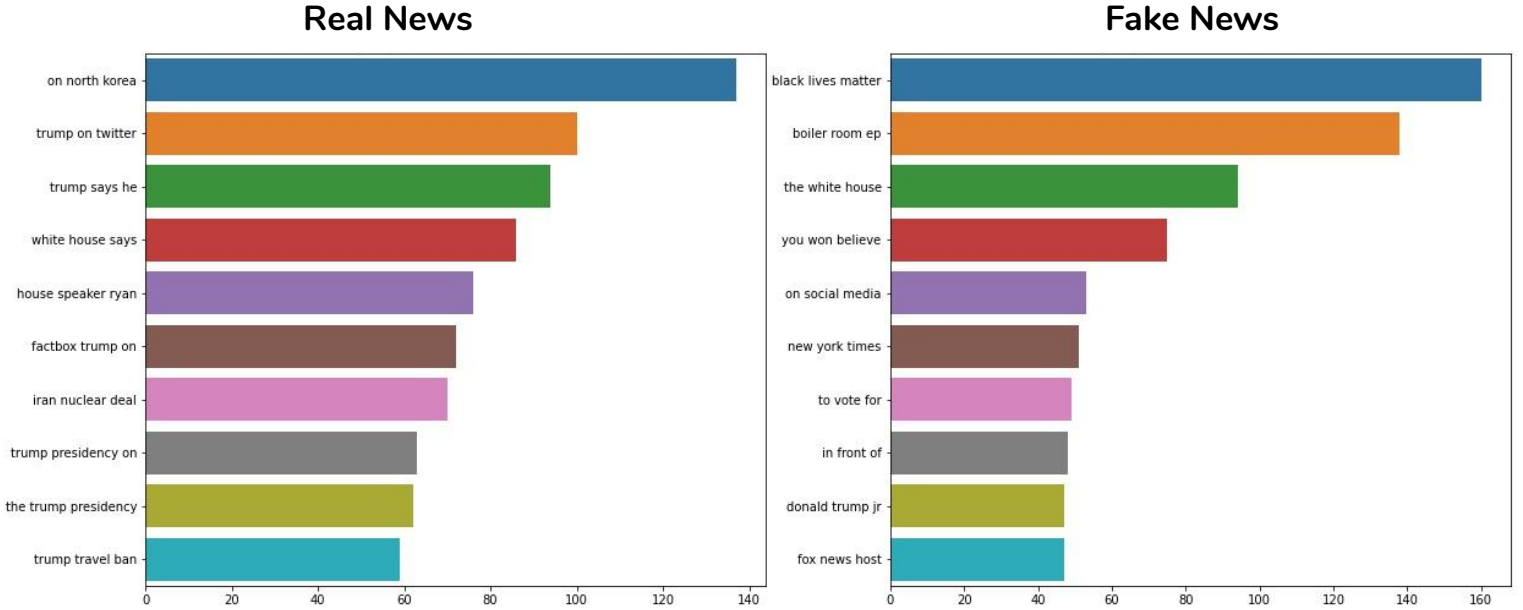
Fake News



Interesting patterns: “trump” is most popular for both real and fake; fake news focuses on relevant people, while real news focuses on relevant issues

Exploratory Data Analysis

Most popular trigrams (word triples) in news article headlines

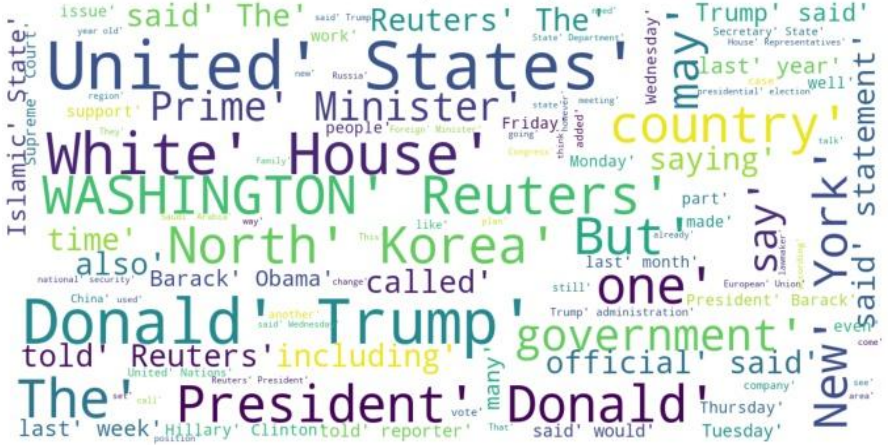


Interesting patterns: topics begin to emerge, specifically that real news focuses heavily around trump (his twitter, presidency, travel ban); fake news focuses on black lives matter, the white house, social media, and fox news

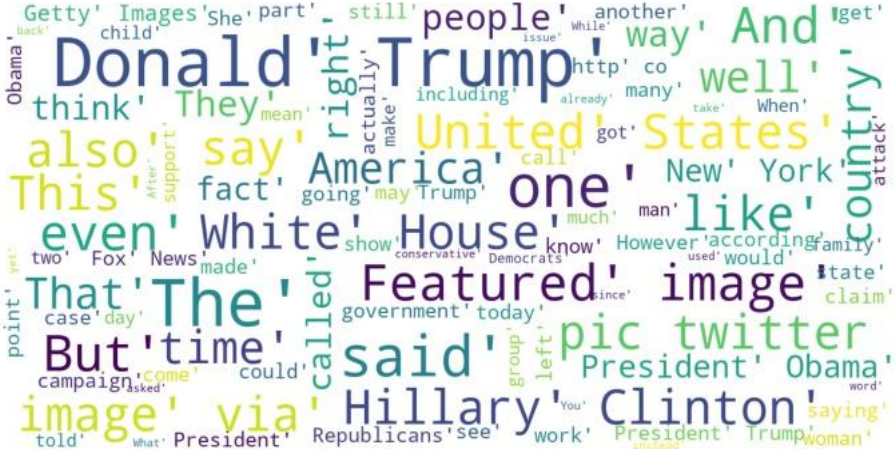
Exploratory Data Analysis

Wordclouds of news article text

Real news



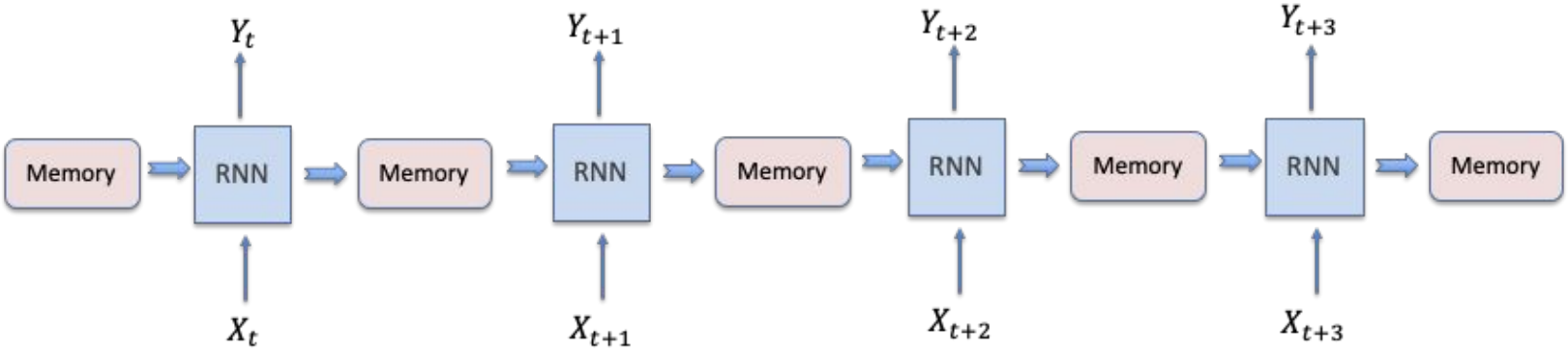
Fake news



Interesting patterns:

- Both types of articles heavily feature Donald Trump, the United States, the White House
- Real news: more focus on international issues, e.g. Islamic State, North Korea, Prime Minister
- Fake news: more focus on Hillary Clinton, Twitter, President Obama

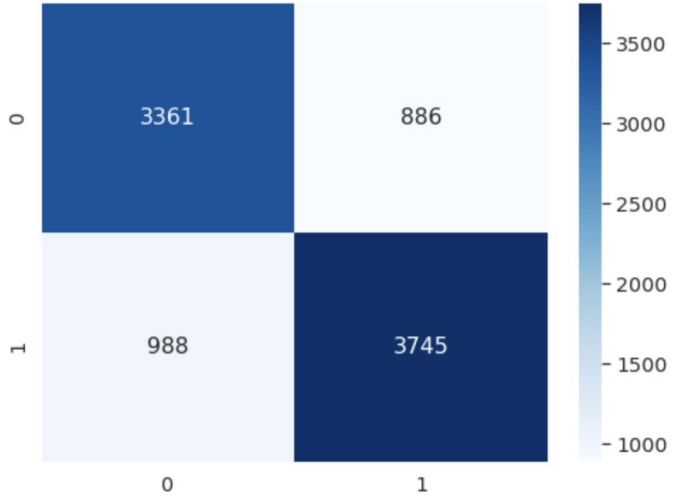
Baseline Model - RNN



➔ Recurrent neural networks recognize the data's sequential characteristics and use patterns to address classification and regression tasks.

Baseline Model - RNN

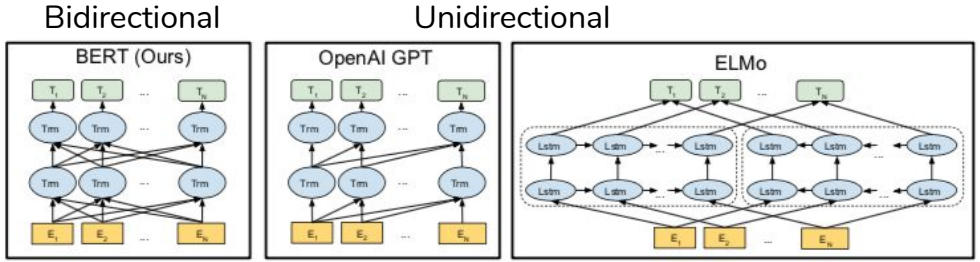
	precision	recall	f1-score	support
0	0.77	0.79	0.78	4247
1	0.81	0.79	0.80	4733
accuracy			0.79	8980
macro avg	0.79	0.79	0.79	8980
weighted avg	0.79	0.79	0.79	8980



Fairly low test accuracy → Room for improvement!

Transfer Learning - BERT

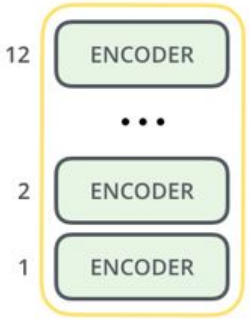
➔ **B**idirectional **E**ncoder **R**epresentations from **T**ransformers



1

➔ Benefits:

- Better contextualized word embeddings
- Pre-trained on so don't need huge amount of data
- Can achieve state of the art results with minimal fine tuning



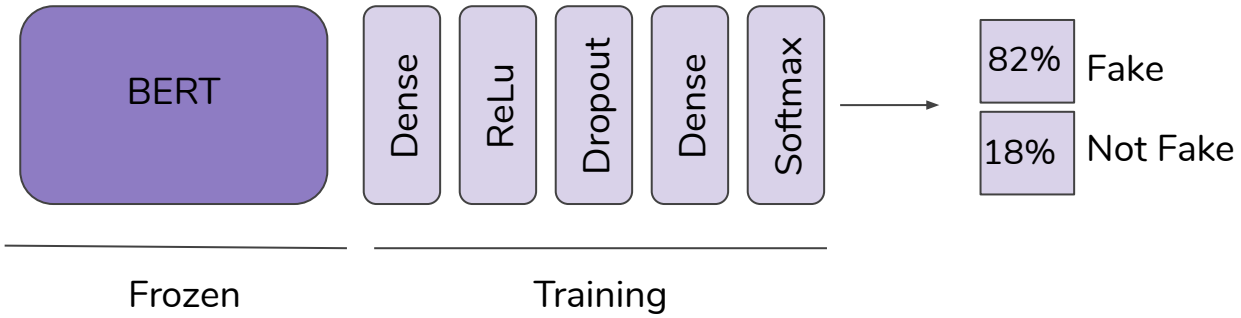
2 BERT$_{BASE}$

1 Devlin et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding 2 http://jalammar.github.io/illustrated-bert/

Transfer Learning - BERT

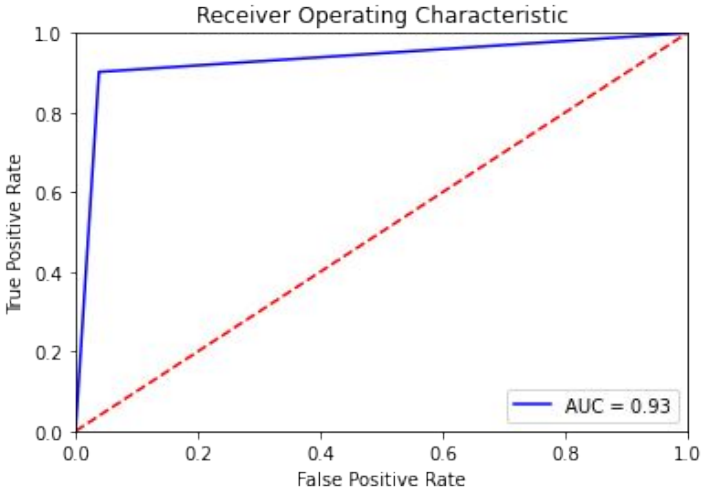
➔ Froze all layers of BERT

➔ Added two additional layers and trained on our data



Transfer Learning - BERT

- ➔ Trained over 5 epochs
- ➔ Final AUC score on test set: 0.9319

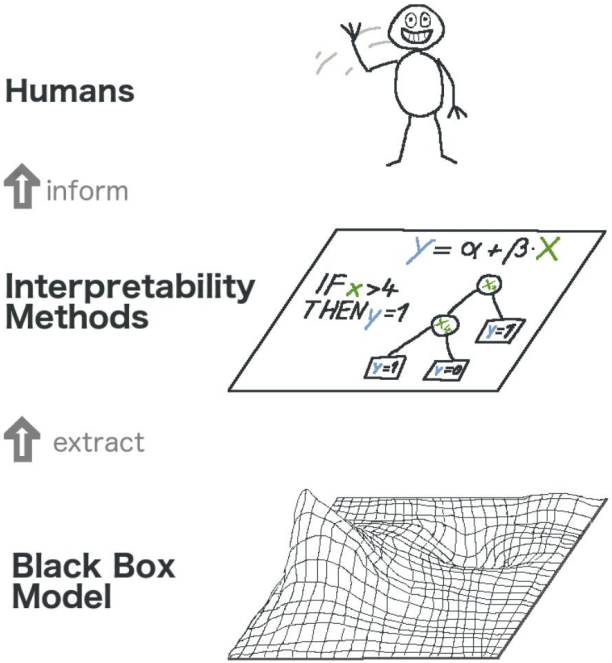


Example of correct prediction (Real News):
Trump names additional senior White House aides: statement. US Republican President-elect Donald Trump on Wednesday filled out his incoming White House senior staff ...

Example of incorrect prediction (predicted as Fake News):
Senate Judiciary panel chief urges FBI to disclose any Trump probes. Senate Judiciary Committee Chairman Chuck Grassley on Thursday called on the FBI to publicly confirm whether or not it is probing President Donald Trump ...

Model Interpretability - LIME

- ➔ Local Interpretable Model-Agnostic Explanations (LIME)
- ➔ Understand how the model works on a case-by-case level
- ➔ Evaluate the trustworthiness of the model



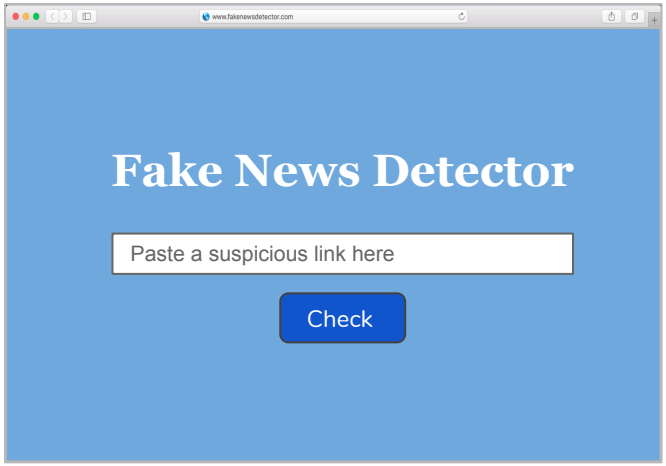
Model Interpretability - LIME

Correctly classified as Real News	Incorrectly classified as Fake news
Trump names additional senior White House aides: statement. US Republican President-elect Donald Trump on Wednesday filled out his incoming White House senior staff jobs, naming three deputy chiefs of staff to help with operations, Trump's transition team said in a statement.	Senate Judiciary panel chief urges FBI to disclose any Trump probes. Senate Judiciary Committee Chairman Chuck Grassley on Thursday called on the FBI to publicly confirm whether or not it is probing President Donald Trump, who said in a letter this week he had been assured that he was not under FBI investigation.
Real	Fake

➔ Expose potential bias and help improving the model

➔ Inform human fact-checkers and accelerate decision making

Conclusions & Future work



Applications demo

Conclusion

Fake news detector via natural language processing

Limitations

- Inherent bias due to limited source of fake news
  - Political tweets
- Ambiguous definition of “fake” news

Future work

- Include diverse news sources on the training set
- Expand the scope of research in other fields (healthcare, environmental science etc.) and cross validate current model
- Launch web application with LIME explanation

Thank You

Team 39

Iris Yoon, Postdoc Applied Mathematics, University of Pennsylvania
Rabiya Noori, MSc Biomedical Engineering, University of Toronto
Jerri Zhang, MS Health Data Science, Harvard University
Renee G. Reynolds, Master of Financial Engineering, UC Berkeley
Hannah Mei, BA Mathematics and Economics, NYU

Acknowledgements

Ronnie Ghose, Tech Lead/Senior SWE, LinkedIn, DS4A Mentor
Yaqi Yang, Data Scientist, Instacart, DS4A Mentor
Savannah Thais, Postdoc, Princeton University, DS4A TA
Correlation One Team

Appendix

Appendix - External validation

Data: 10,000 fake news articles from 2016 (Kaggle)

	precision	recall	f1-score	support
0	0.00	0.00	0.00	0
1	1.00	0.82	0.90	9482
accuracy			0.82	9482
macro avg	0.50	0.41	0.45	9482
weighted avg	1.00	0.82	0.90	9482

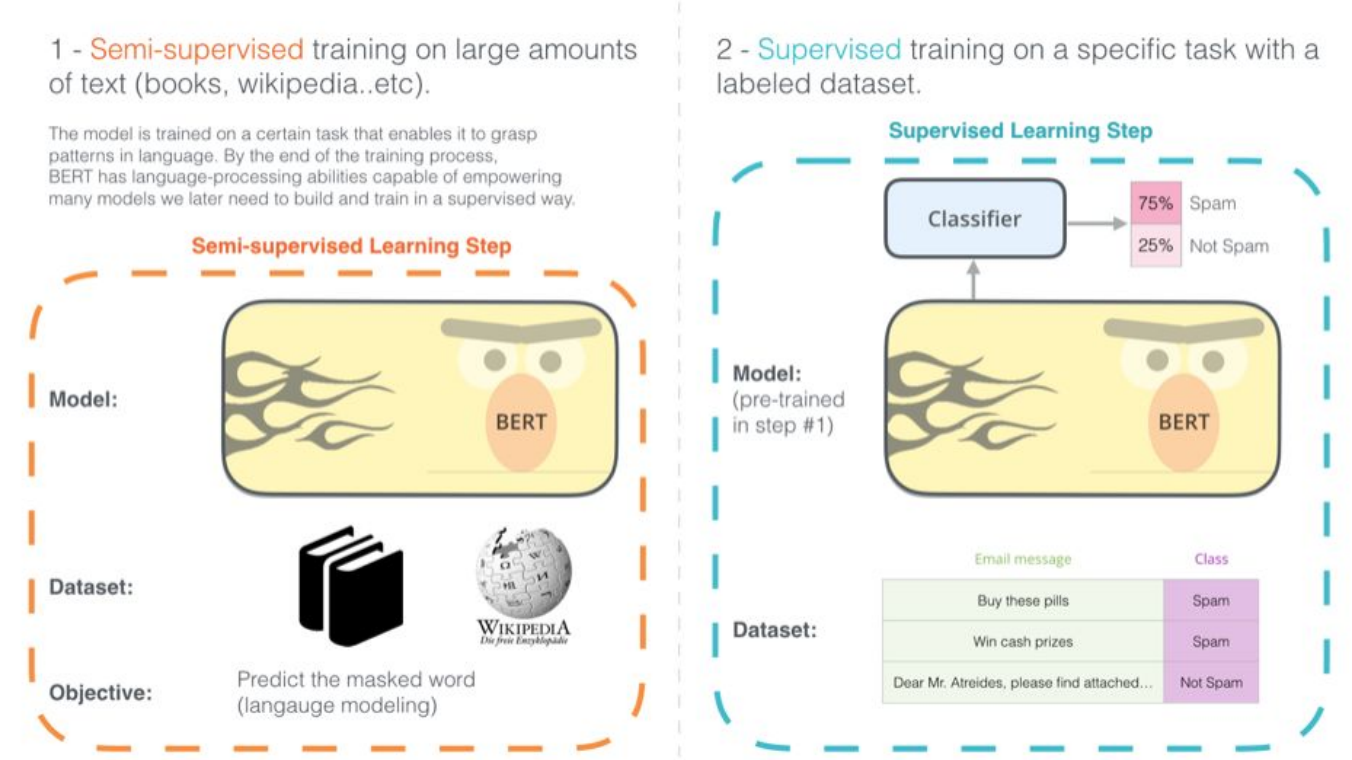
Decrease in performance

External validation set contains more diverse fake news

- Non-political (fake science)
- Biased news
- Hate news

BERT Details - pre-training

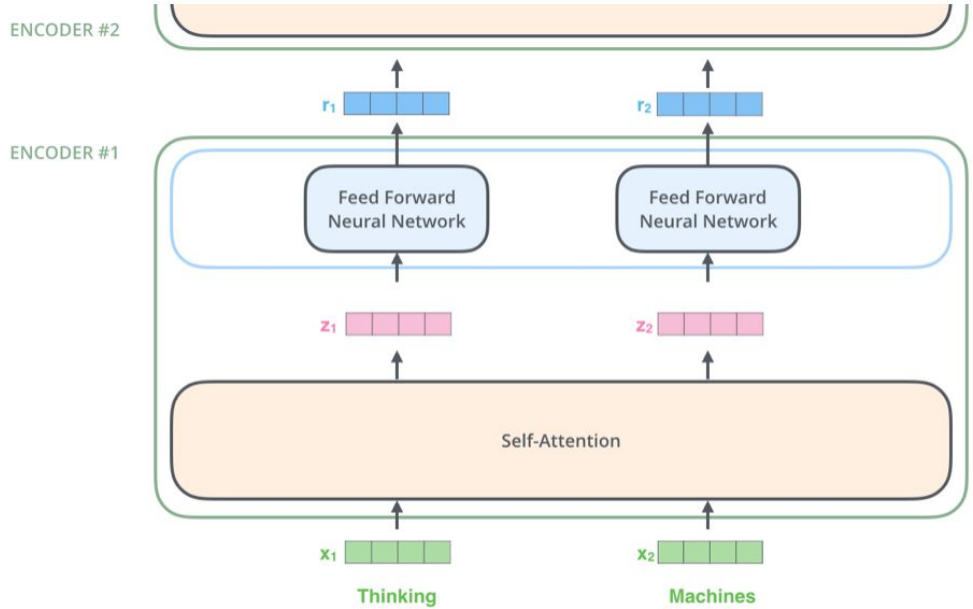
http://jalammar.github.io/illustrated-bert/



The two steps of how BERT is developed. You can download the model pre-trained in step 1 (trained on un-annotated data), and only worry about fine-tuning it for step 2. [Source for book icon].

BERT Details - Transformers architecture

https://jalammar.github.io/illustrated-transformer/



The word at each position passes through a self-attention process. Then, they each pass through a feed-forward neural network -- the exact same network with each vector flowing through it separately.