

CON

TEXTUAL

Uncovering news context through transfer learning

MSDS2020 TeamTextual

**back in 2016,
The US saw a battle
for the ages:**



Over the course of this election, one of the hottest topics in the news cycle was **"fake news"**.



Don't blame the election on fake news. Blame it on the media.



ANALYSIS

By **Duncan J. Watts**
and **David M. Rothschild**

DECEMBER 5, 2017

Photo by Joe Raedle/AFP/Getty Images

SINCE THE 2016 PRESIDENTIAL ELECTION, an increasingly familiar narrative has emerged concerning the unexpected victory of Donald Trump. **Fake news**, much of

CALL FOR SUBMISSIONS

2020 ROBERT F. KENNEDY
BOOK & JOURNALISM AWARDS

Media bias against conservatives is real, and part of the reason no one trusts the news now

It might not be conscious, but the way that reporters treat conservatives in their coverage has always shown their liberal leanings





Maria Ressa ✓

@mariaressa



#Presstitute is part of a global attack against journalists. First used in PH by one of President Duterte's anchor bloggers, it's been used to attack journalists in the US, India, South Africa - along with the march of cheap armies on social media.



Lynzy Billing ✓ @LynzyBilling · Feb 28

Replying to @mariaressa and @washingtonpost

I'm seeing a pattern: 28 tweets with the word "presstitute"...

example of a journo practicing presstitute...

replying to @LynzyBilling @washingtonpost and @RegineCabato

Only fools would beleive this lazy ass bitch

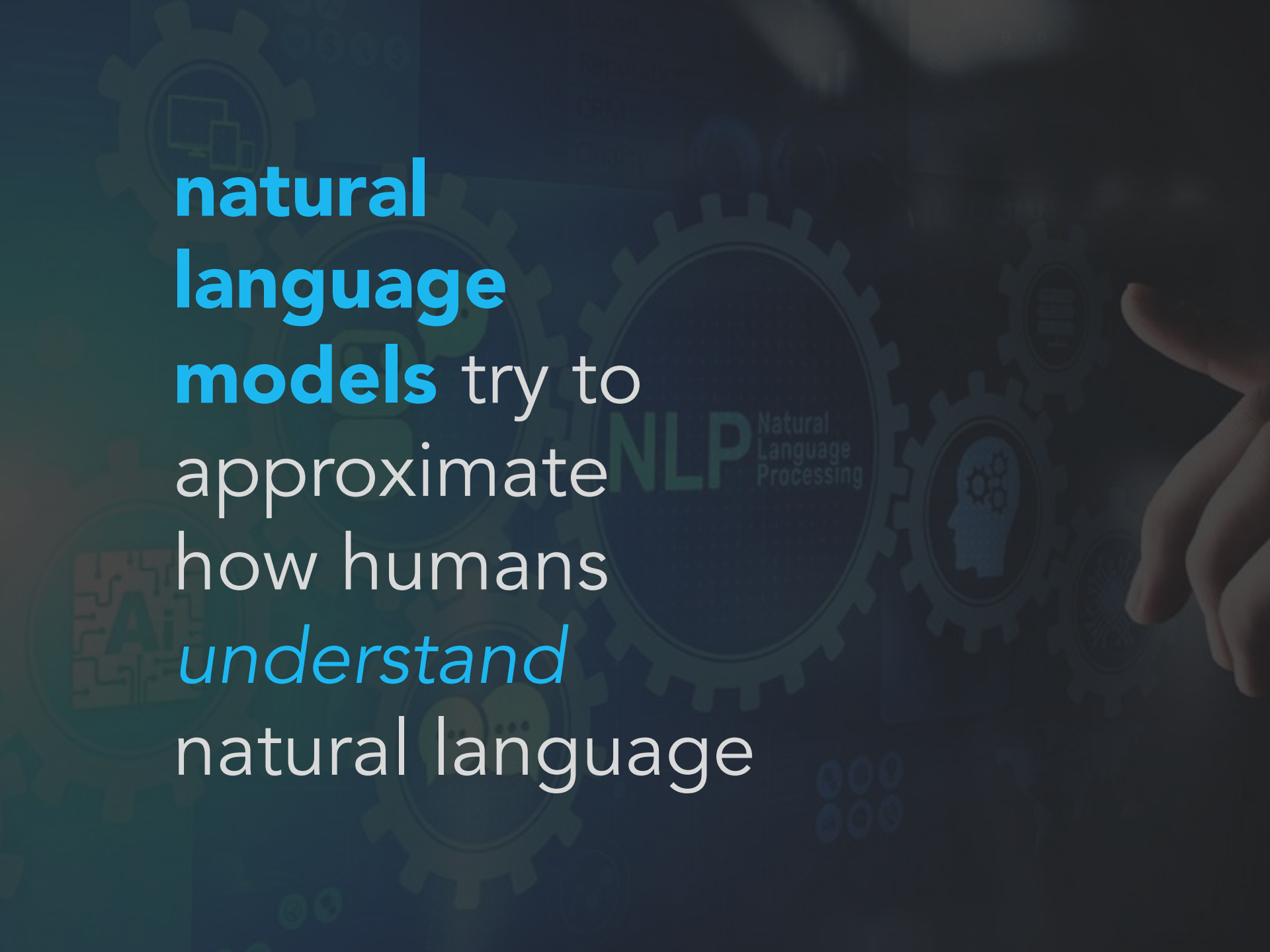
We ask ourselves:

what is the
effect of
media bias
to **human**
perception?



and what has
data science
got to do
with it?



The background is a dark blue gradient with several faint, semi-transparent icons. On the left, there are gears and a bar chart. In the center, a large gear contains the text 'NLP Natural Language Processing'. To the right, there is a silhouette of a human head with gears inside, and a hand pointing towards the center. At the bottom right, there are some small, faint icons resembling a grid or data points.

**natural
language
models** try to
approximate
how humans
understand
natural language

Introducing GPT

Generative

Pretrained

Transformer

GPT 2

large-scale
unsupervised
language model



OpenAI

who?

WIRED

BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY TRANSPORTATION

SIGN IN

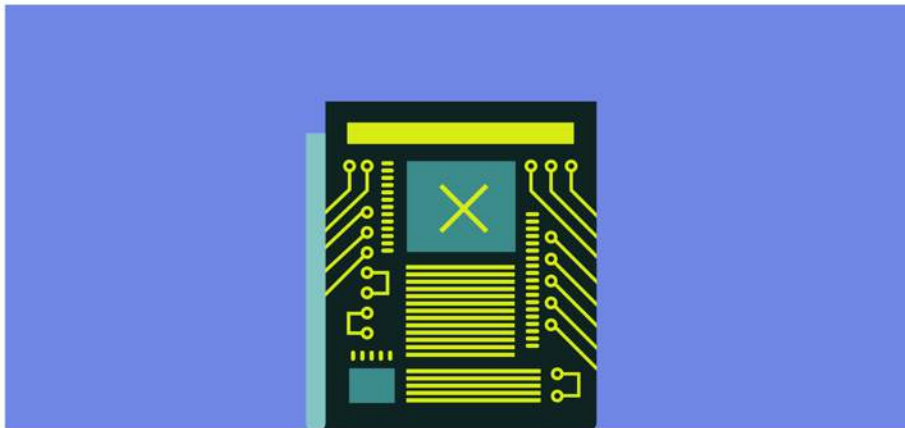
SUBSCRIBE

TOM SIMONITE

BUSINESS 02.14.2019 12:08 PM

The AI Text Generator That's Too Dangerous to Make Public

Researchers at OpenAI decided that a system that scores well at understanding language could too easily be manipulated for malicious intent.



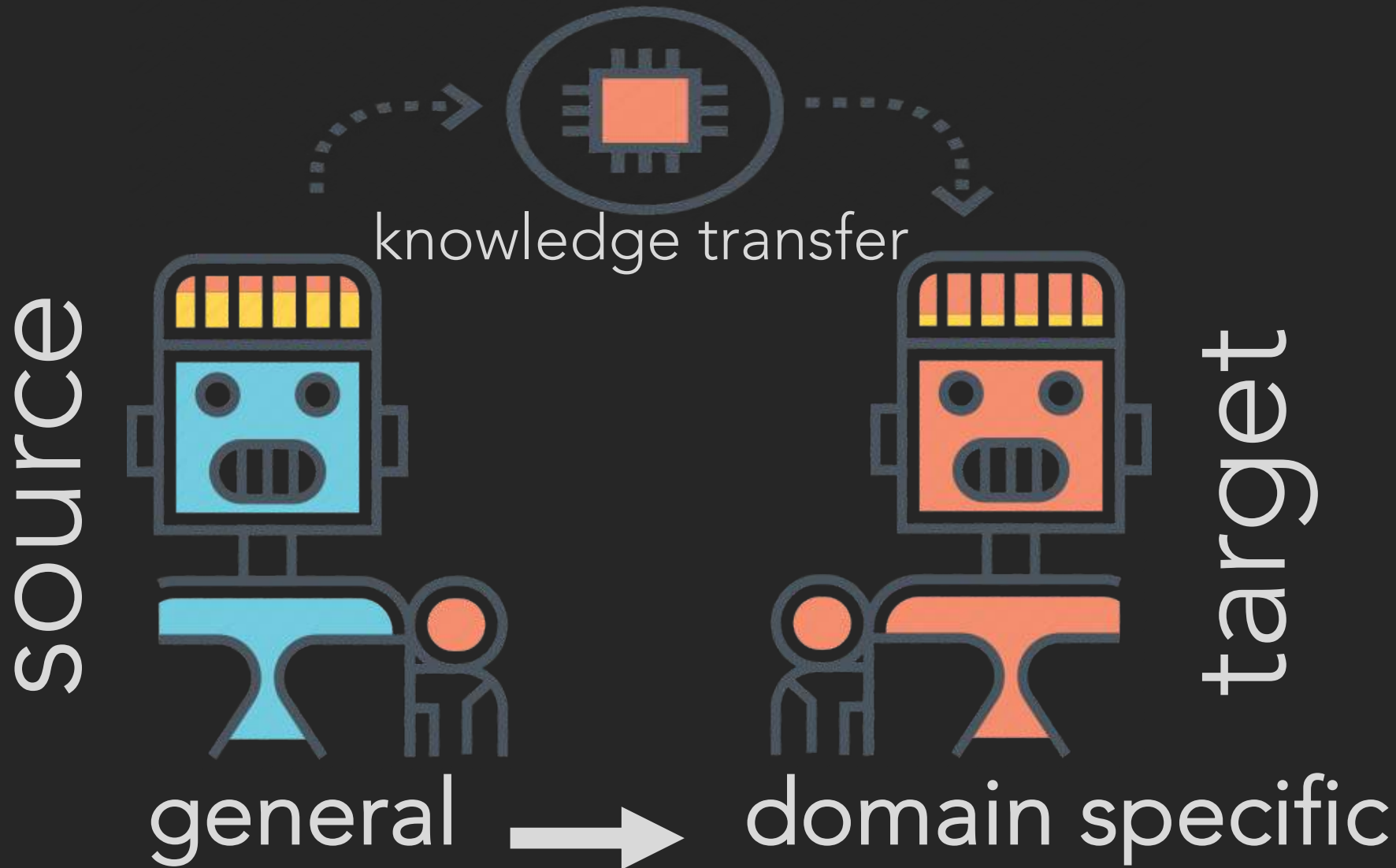
GPT

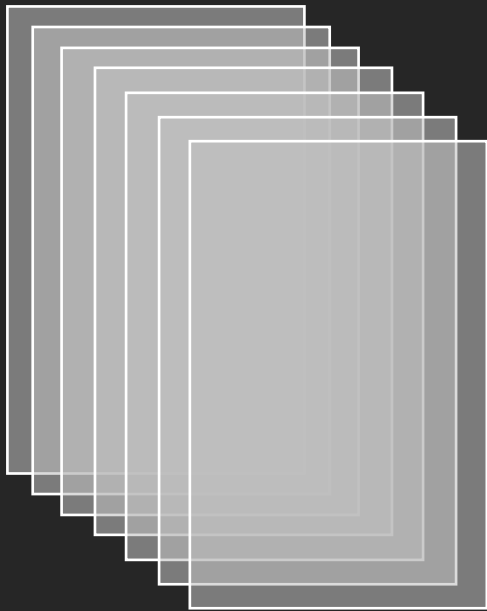
predict the next
word, given all of
the previous
words within
some text

Can we train GPT2 to
understand and,
more importantly,
generate context
based on articles
from different news
outlets?



what is **transfer learning**?





News Dataset

The New York Times



wop



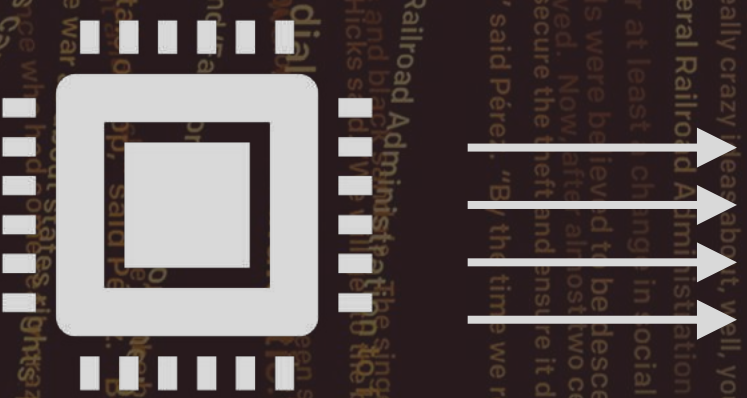
The Atlantic

18 American Publications
200,000+ articles
2015-2017



We train:
18 models
~355M
Parameters

GPT 2



But first...

let's define some terms

Polarity

How positive or negative is the overall sentiment?

Positive sentiment

"I love this cake, it's the best!"

Negative sentiment

"This is the worst cake I've ever eaten. I hate it."

Subjectivity

How objective or subjective is the overall text?

Object sentiment

"This cake has chocolate."

Negative sentiment

"This cake tastes very good!"

The Results

We fed our trained models, one for each major news outlet in the US, and asked them to complete the following prompt:



“Donald Trump...”



“Hillary Clinton...”

The Results

Disclaimer: These results are all completely generated by the re-trained GPT-2 Model, and no human involved in creating these fake news articles.

The Atlantic



Donald Trump's #1 target: the undocumented, and the millions of undocumented immigrants who are already here, including children, will be here for at least the next four decades., These are the people the president wants to deport: mothers who care about their children, elderly people who keep a home for their grandchildren, and people who have committed no crime.

BUSINESS INSIDER



Hillary Clinton-aligned groups, the Democratic Party has an opportunity to take a significant step toward rebranding itself as the party of the working class, the party of the underclass, and the party of immigrants. Breitbart, the far-right news website known for its tough-on-crime stance, is often associated with the alt-right, a far-right movement that advocates for the use of harsh immigration policies to combat illegal immigration.

A look at subjectivity and polarity in media.

We look at the performance of our trained model for each news outlet with regard to all the articles for Trump and Clinton.

In this sense, we can say that the learned representations of the model are a snapshot of the overall news outlet.

polarity



subjectivity

polarity



BUSINESS INSIDER

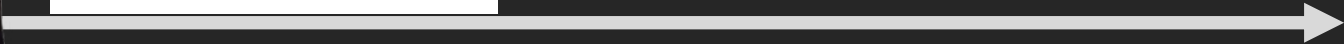


The New York Times

The Atlantic



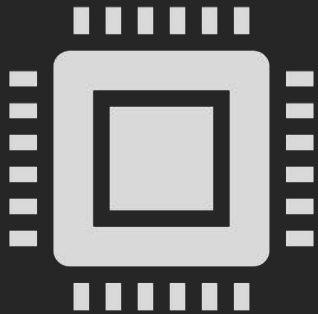
subjectivity



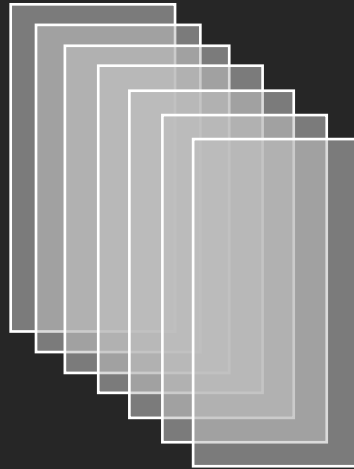
BREITBART



GPT2



News Dataset



Transfer Learning

Media Bias

**Thank
you!**