



# CISPA

HELMHOLTZ CENTER FOR  
INFORMATION SECURITY

## Regulating disinformation campaigns with artificial intelligence

Ninja Marnau



*Disinformation:  
“false information deliberately  
and often covertly spread  
(as by the planting of rumors)  
in order to influence public opinion  
or obscure the truth”*



# 'I remember Pearl Harbor': Inside Trump's hot-and-cold relationships with Japan's prime minister

**BREITBART**



TRENDING: TRUMP INAUGURATION SUBPOENAED CARAVAN MARCH TO WALL-LESS TX NORTHAM: I'M S

## SOME CARAVAN MIGRANTS HAVE HIV/AIDS AND TB, SAY MEXICAN AUTHORITIES

SHARE 62,076 EMAIL SHARE TWEET



File Photo: PEDRO PARDO/AFP/Getty Images

by BOB PRICE | 29 Nov 2018 | 12,722



**Brittany Pettibone** @BrittPettibone

#PizzaGate is a world-wide citizen investigation now. It cannot be stopped. At this point, the truth being brought to light is inevitable.

RETWEETS 5,297 LIKES 6,467

3:01 AM - 20 Nov 2016

Nem os mortos respeitam!!!

**Thiago Nunez Souza** Ontem às 18:45

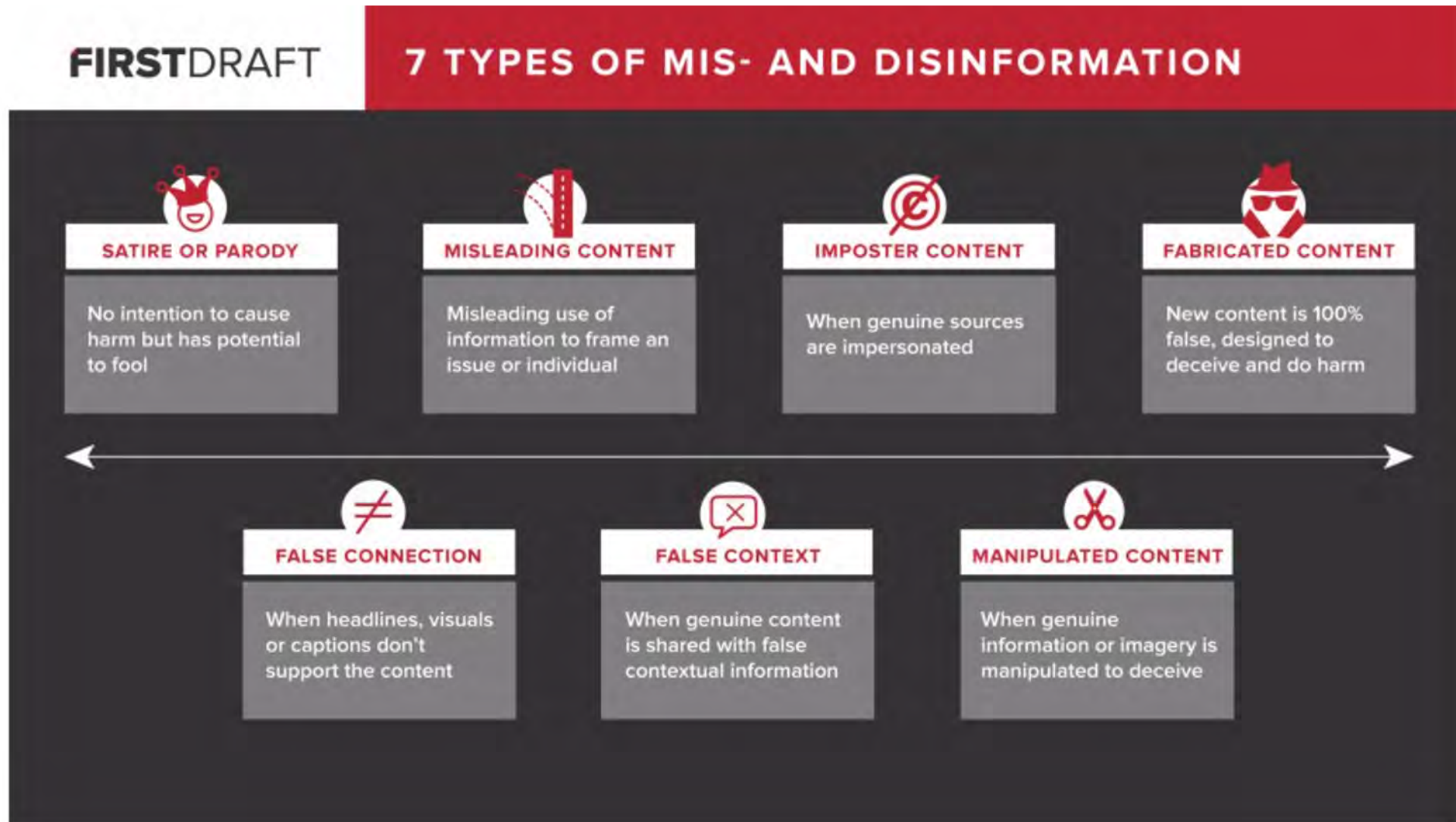
ESTA SENHORA FOI AGREDIDA POR PETISTAS NA RUA QUANDO GRITOU BOLSONARO.

Curtir Comentar Compartilhar

191

5.101 compartilhamentos

# What is Fake News? What is Disinformation?



Source: Claire Wardle of FirstDraftNews.com

# What motivates spreading of disinformation?

- Insufficient research
- Political agenda
- Misinformed beliefs
- Undermining trust in authorities and processes
- Distraction of attention
- Conspiracy theories
- Display purposes
- Provocation



# Marketplaces Selling Tools and Services for Public Opinion Manipulation Campaigns

Manufacturing fake news requires tools, and the online underground is rife with them. In principle, some of the services we identified can also be used legitimately, such as in content marketing, but like any potent tool we see them discussed on underground forums, abused, and leveraged to disseminate fake news and manipulate public opinion.

- Engagement economy
  - Algorithmic boost to content that gets user engagement
  - Users engage when the content triggers emotions
  - Disinformation & fake news trigger (mostly negative) emotions and thus engagement

## Microtargeting

If you were ...

Someone who "likes" patriotism



Heart of Texas

Sponsored · 🌐

👍 Like Page

Border Patrol agents in South Texas arrested an illegal alien from Honduras that had previously been deported and convicted of Rape Second Degree.

Thanks to Obama's and Hillary's policy, illegals come here because they wait for amnesty promised. The wrong course had been chosen by the American government, but all those politicians are too far from the border to see who actually sneaks through it illegally.

Rapists, drug dealers, human traffickers, and others. The percent of innocent poor families searching for a better life is too small to become an argument for amnesty and Texas warm welcome.



3.1K Reactions 89 Comments 1.2K Shares

👍 Like    💬 Comment    ➦ Share

If you were ...

A Bernie Sanders supporter



Born Liberal

Sponsored · 🌐

👍 Like Page

"If you ask me about the Clinton Foundation, do I have a problem when a sitting secretary of state and a foundation run by her husband collects many millions of dollars from foreign governments, many governments which are dictatorships... yeah, I do," Sanders said in an interview with Jake Tapper on CNN.

So, Bernie has the point. I appreciate the way he maintains his independent opinion in spite of all the pressure that is on him as the "second Democratic option". Because that ... [See More](#)



124 Reactions 5 Comments 21 Shares

👍 Like    💬 Comment    ➦ Share



## Deepfakes:

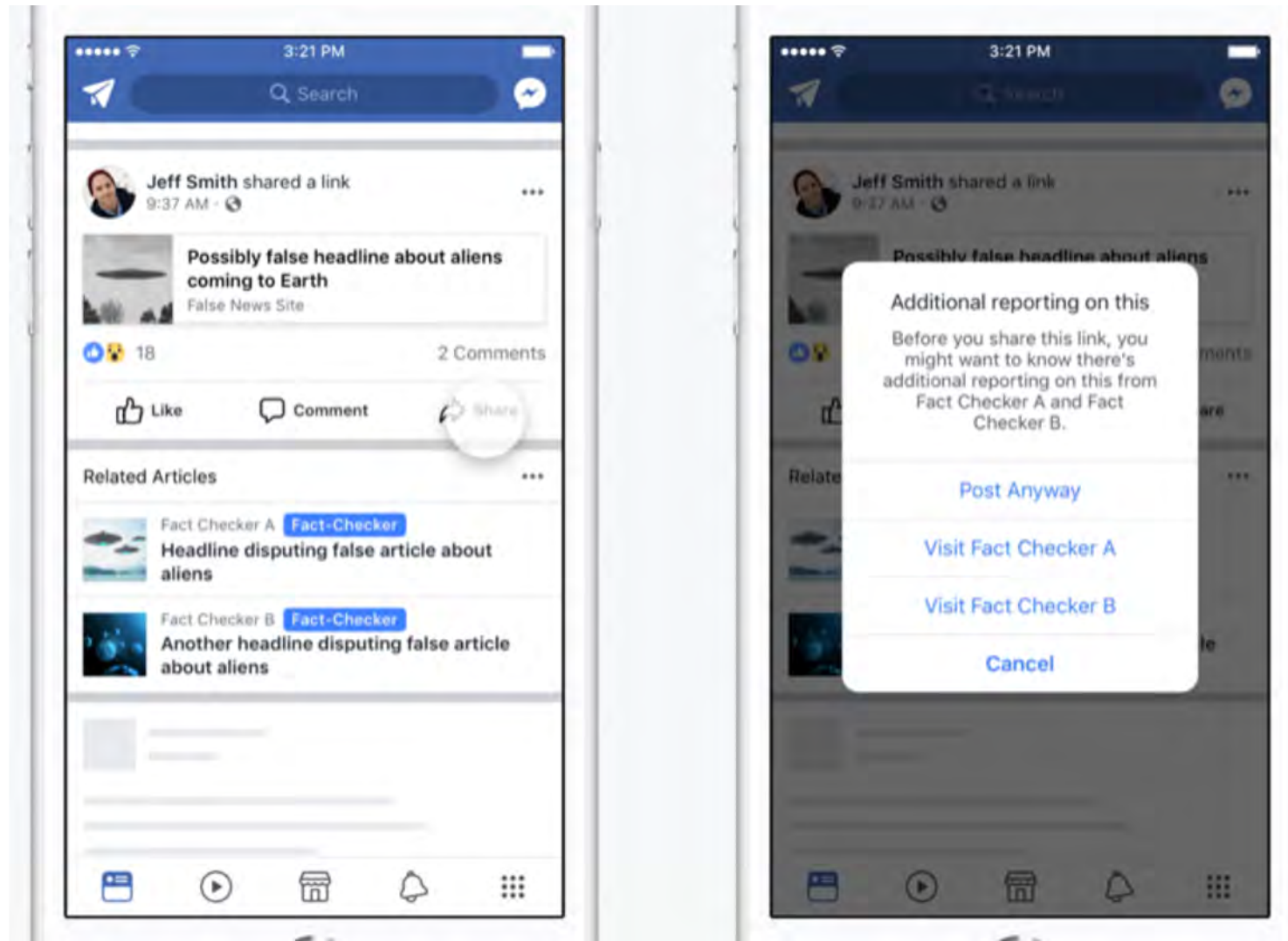


**AI for marking of disputed content**

**AI for bot detection**

**AI for fact checking**

## AI for marking of disputed content





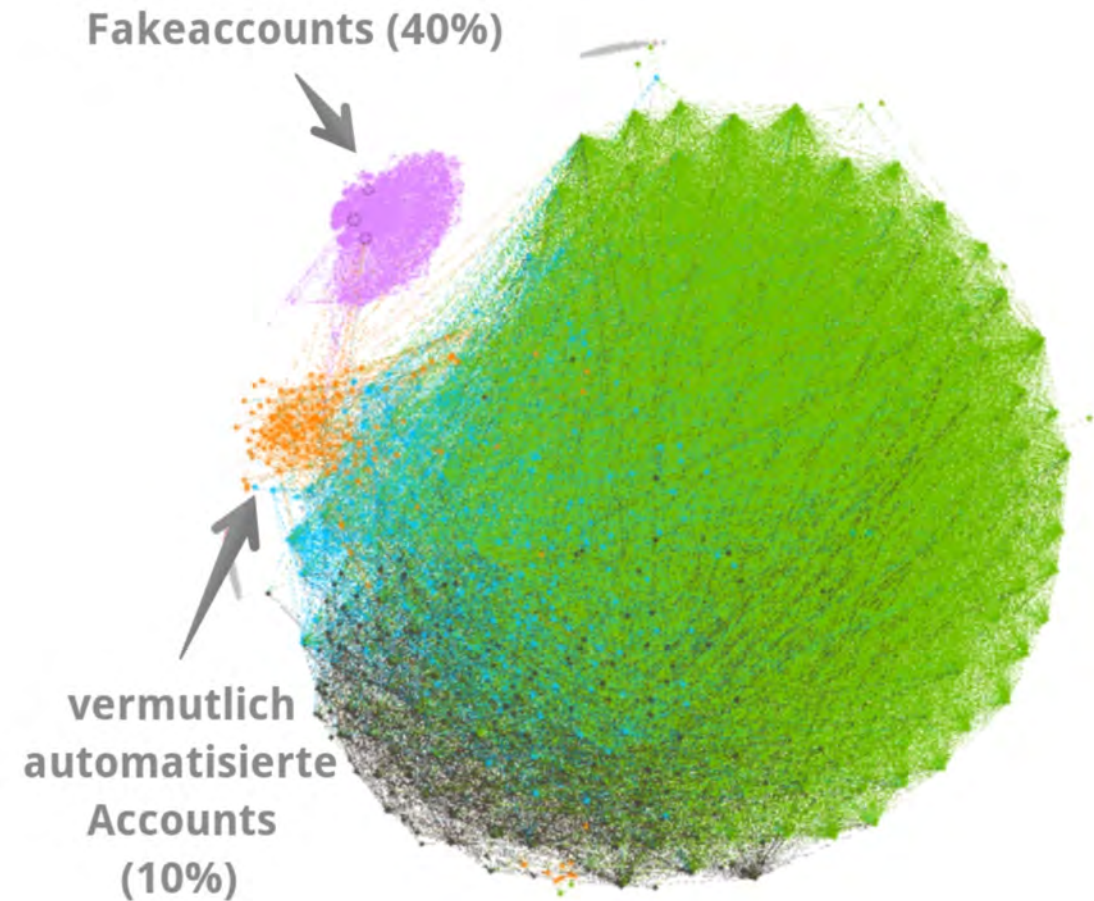
## AI for bot detection

- Availability of Social Network data is severely limited
  - Very few data points from e.g. WhatsApp, Facebook, almost all research is done on Twitter
- Ground truth
  - How to be sure an account is a bot?
- Not all bots are bad



Conclusion: Machine learning with manual confirmation probably best

Bot detection –  
mostly a phantom problem



Source: Luca Hammer (analyst)

## AI for fact checking

Combining Machine Learning and Natural language processing

Mostly unsolved challenges:

- Detecting irony/satire/comedy
- Identifying quotes and critique
- Complicated negations



# CredEye

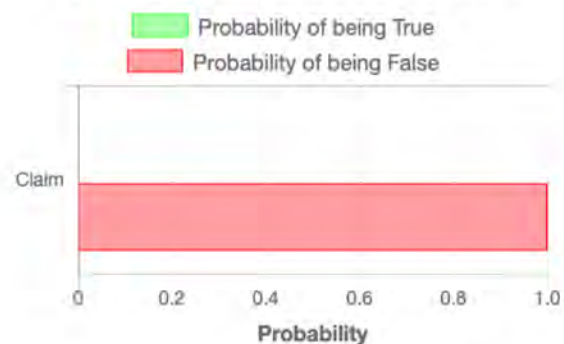
A CREDIBILITY LENS FOR ANALYZING AND EXPLAINING MISINFORMATION

Claim

Examples:

Entire Web
  All frames
  US News
  UK News
  Social Media

## Credibility Assessment



Do you agree with the assessment?



## Evidence

Show Details:

*Barack Obama* · Born Aug 04, 1961 · Honolulu, HI *Barack Obama*. Obama was born in 1961 in Honolulu, Hawaii, two years after the territory was admitted to the Union as the 50th state. Raised largely in Hawaii, he also lived for a year of his childhood in the State of Washington and four years in Indonesia. After graduating from Columbia University in 1983,...

*Barack Obama* · Wikipedia  
[en.wikipedia.org/wiki/Barack\\_Obama](http://en.wikipedia.org/wiki/Barack_Obama)  
 — bing.com

Credibility Score: 0 Web-source Trustworthiness: 1 Article Stance (Refute Score): 1

\*Word also present in the claim      \*Top Language Features of True and False Claims

\*Credibility Score -- Probability of claim being true, calculated using language and stance related features of the article

\*Web-source Trustworthiness -- Trustworthiness of the web-source in the range of 0 (least trustworthy) to 1 (most trustworthy)

# CredEye

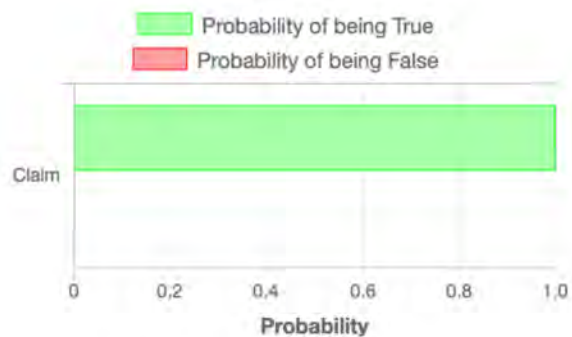
A CREDIBILITY LENS FOR ANALYZING AND EXPLAINING MISINFORMATION

Claim:

Examples:

Entire Web
  All News
  US News
  UK News
  All 120000

## Credibility Assessment



Do you agree with the assessment?



## Evidence

Show Details:

*In a campaign speech in Ft. Lauderdale, Florida, Donald Trump called President **Obama** the 'founder of Isis' and Hillary Clinton the group's 'co-founder'. Update: After two days of insisting he was serious and literally meant that Barack **Obama** founded ISIS, Donald Trump tweeted on 12 August 2016 that he was only being sarcastic: Ratings challenged @CNN reports so seriously that I call President **Obama** (and Clinton) "the founder" of ISIS, & MVP. Donald Trump: 'Obama Founded ISIS' - snopes.com www.snopes.com/news/2016/08/11/obama-founded-isis/ Is this answer helpful? —bing.com*

Credibility Score: 1 Web-source Trustworthiness: 1 Article Stance (Refute Score): 0

\*Word also present in the claim and **False** Claims

\*Credibility Score -- Probability of claim being true, calculated using language and stance related features of the article

\*Web-source Trustworthiness -- Trustworthiness of the web-source in the

- Prevalence:
  - 90% of human users don't share or interact with fake news (but might see it)
  - Very few accounts (mostly human-operated) act as amplifiers
  - Bots and automated accounts are only small part of the problem
- Regulation needs to be aimed at clearly defined types of misinformation to protect free speech and freedom of press
- Dark social, microtargeting & deep fakes will amplify the problem
- AI solutions are not a silver bullet
- “Homegrown desinformation” - politicians and media outlets that use desinformation



