
The impact of Artificial Intelligence on social media

Lucila Moreno & Carolina Rodriguez - 11 de octubre de 2018

Prof. German Rigau

Abstract

This report presents a brief analysis of the problematic of fake news, the negative impact it has on many countries and how social networks try to fight against it, knowing that it is a long-term battle that will get worse over the years as technology will get more sophisticated at mimicking human behavior.

Introduction

Social networks nowadays play an important role in society. People have found a media where they can express themselves without censorship. Although mostly everyone makes a positive use of social networks, others abuse of the freedom of expression. Although the term “fake news” has existed for a long time, in the past years has gained everyones attention, due to its appearance on social medias where this type of disinformation is easily and quickly shared. Society has reached a point where people rely on any post in spite of the author and this disinformation is causing a significant number of deaths everyday. Social networks´ leaders are constantly trying to fight against this problematic and opt to start using artificial intelligence to tackle it.

On the other hand, there is a rising tendency to create misleading news with artificial intelligence, which involves a bigger problem for companies to detect, due to its approximation to human-quality handwritten news.

Description of the problematic

We live in the post-truth era, where true and false are confused in such a way that knowing reality seems like an almost impossible mission. History, both past and recent, is full of rumours or even falsehoods that are spread as truths and then have drastic consequences, since the German Parliament's fire in 1933, which Hitler manipulated in his favor by attributing it to the Communists. However, todays technology allows a great paradox; while the access to information is democratized, the false news lives its own boom. Thus, more information would not necessarily lead to more truth, but to the proliferation of rumors that turn into false information.

The 'fake news' made its stellar appearance during the election campaign that brought Donald Trump to the White House. According to Facebook, only 1% of the publications on its platform are false, but that contrasts sharply with the recognition it has had to make before the US Congress on the impact that Russian propaganda had on the 2016 election.

A survey conducted in the US by the Edelman agency [1] tries to clarify some points on the matter. Its main conclusion is that most Americans believe that the media create false news with a single goal: the destruction of political opponents. According to the data, false news is decreasing confidence in traditional news companies.

Also researchers from the Media Lab of the Massachusetts Institute of Technology have carried out a study that has shown that false news has a disturbing power, because it reaches many more people than real information and can even alter the criteria to distinguish between what is true and what is false. According to a study on the impact of fake news in Spain, 86% of Spaniards do not easily distinguish invented news from real news.

But how does the ordinary citizen understand what the 'fake news' are? For 47% of those surveyed by Edelman, the false news is "careless or biased texts of the news organizations", while another 39% ventures to describe them as an "insult that is used in excess to discredit the news that people do not like them." For each person who recognized Twitter as a technology company, there were five that identified it as a news portal. More than half of the respondents also accuse these platforms of spreading false news in the world.

On the other hand, false news is changing consumer habits when reading or accessing the media. More than half of Americans (51%) say they are now more careful about where they consume information, as well as using more verification tools than ever before.

"We have locked ourselves into a cave of our own making. We have become tribal and defensive. The mainstream media is categorized as elitist and politicized, unable to carry out its vital role of truth-telling as the fourth estate in global governance. We opt instead for opinion based on personal experience, using the social platforms as equivalent or superior forms of communication." concludes Richard Edelman.

Nowadays, about 100 million pieces of information hit the internet, this massive and uncontrolled amount oftentimes lead to systematic spread of inaccurate and misleading information on the Web and social media poses a major risk to society.

Is worth mentioning that digital disinformation is thriving on assortment of cognitive, social, and algorithmic biases and most important the incapacity of the current countermeasures based on journalistic corrections do not seem to scale up.

According to a research published by PNAS (Proceedings of the National Academy of Sciences)[2], the digital misinformation posted on social media is one of the main threats to our society. This five year study, that investigates how rumours and lies go viral on the web, looked at a diverse group of Facebook users and found that the platform created 'echo chambers' which allowed people's already held views to be confirmed, and prejudices to be reinforced.

As the website iflscience.com explains in its report on the research: “echo chambers describe certain areas of the media, particularly the internet, wherein information or beliefs are reinforced by repetitive transmission inside an enclosed virtual space. These spaces, which also serve to keep contrasting views at bay, may explain why there are so many groups of people online (particularly on Facebook) that steadfastly believe information that is demonstrably nonsensical.”[3] The report's authors looked specifically at two areas, scientific and conspiracy news. The latter included the case of the US military exercise Jade Helm15 [4], where conspiracy theorists spread stories online that a training exercise across seven states was actually an invasion and military takeover of the country. The stories included statements that some branches of US supermarket Walmart were to be turned into prisons. Furthermore, they stated “Users tend to aggregate in communities of interest, which causes reinforcement and fosters confirmation bias, segregation, and polarization. This comes at the expense of the quality of the information and leads to proliferation of biased narratives fomented by unsubstantiated rumors, mistrust, and paranoia.”

How artificial intelligence is used to detect and create fake news

Fake news is unlikely ever to be fully eradicated, but implementing machine learning tools raises the barrier regarding the expertise and technology needed to deceive.

As simple as it seems, it may be difficult to train an AI to ignore minor infractions and focus on the more nefarious activity. Drawing this line can also expose – and amplify – certain biases in society. This can relate to gender, racial background or neighborhood stereotypes. It can even have political consequences, potentially restricting expression of particular viewpoints. Nowadays, the most effective way is to try to divine the motivation of the media pushes and take the necessary actions.

Focusing the attention on one particular case, fake news in Facebook has been a hot topic on every newspaper and television programmes worldwide due to the negative impact it has been having in countries such as India, Pakistan and particularly Myanmar, where Facebook is almost the only source of information within reach to them causing an excessive trust on every new posted on this social network. [7] Fake news for political, social and economical purposes has caused many deaths in this Southern Asia countries.

For this reason, and many others, Facebook started implementing artificial intelligence to try to detect fake news. A few months ago, Mark Zuckerberg during a conference in San Jose, California revealed the way Facebook makes use of Artificial Intelligence tools, mainly in seven areas: nudity, graphic violence, terrorist content, hate speech, spam, fake accounts and suicide prevention.

In the case of graphic violence and nudity, Facebook uses a software called “computer vision” that detects certain kind of elements on images. On the other side, hate speech is far more complex to detect, as there is no pattern and you have to see the intention of the speech, as it may be sarcastic or self-referential, for example. [6] Moreover, AI tools are not yet prepared to identify every word from the less widely spoken languages. Furthermore, new measures are being applied to fight against disinformation including fact-checking programmes such as Check and Snopes.

The biggest challenge of using artificial intelligence to detect fake news nowadays is that it puts technology in an arms race with itself. Machine learning systems are proved to be capable of creating fake news or even fake images and videos, nowadays called deepfakes which consist of replacing ones face with another one to simulate an action or speech that never occurred in a sophisticated and realistic way.

Aspects to improve in the future

Although experts have been trying to figure out how to combat fake news for decades, with the aparison of social medias, this task has become more complex. Despite the successes reached so far, and although we understand the difficulty of developing an algorithm capable of understanding the intention of certain post and having the necessary capacity to judge whether it should be deleted or not, there is yet a whole world to discover.

As mentioned above, it will take an AI to catch an AI, dueling each other to determine what's real. Researchers are already preparing to use AI to identify these AI-created fakes. For example, techniques for video magnification can detect changes in human pulse that would establish whether a person in a video is real or computer-generated. But both fakers and fake-detectors will get better. Some fakes could become so sophisticated that they become very hard to rebut or dismiss – unlike earlier generations of fakes, which used simple language and made easily refuted claims.

The best way to combat the spread of fake news may be to depend on people. The societal consequences of fake news – greater political polarization, increased partisanship, and eroded trust in mainstream media and government – are significant. If more people knew the stakes were that high, they might be more wary of information, particularly if it is more emotionally based, because that's an effective way to get people's attention.

Bibliography

1. Edelman trust barometer special flash poll- Edelman. November 2017.
Online: <https://www.edelman.com/trust2017/trust-barometer-media-fake-news-flash-poll/>
2. The spreading of misinformation online- Proceedings of the National Academy of Sciences. December 2015.
Online: <http://www.pnas.org/content/pnas/early/2016/01/02/1517441113.full.pdf>
3. How Misinformation Spreads On The Internet - Robin Andrews. January 2016.
Online: <https://www.iflscience.com/technology/facebook-echo-chambers-help-spread-and-reinforce-misinformation/>
4. Operation Jade Helm: why conspiracy theorists fear the US is invading Texas - Adam Gabbatt (The Guardian). July 2015.
Online: <https://www.theguardian.com/us-news/2015/jul/15/jade-helm-texas-conspiracy-theory>
5. A multi-dimensional approach to disinformation- High Level Expert Group of the European Commission. March 2018.
Online: <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>
6. How Facebook uses artificial intelligence to take down abusive posts - Cnet. May 2018.
Online: <https://www.cnet.com/news/heres-how-facebook-uses-artificial-intelligence-to-take-down-abusive-posts-f8/>

Annex

7. To have a perspective of the great impact that the propagation of false information can generate, it is worth mentioning the crisis that the Rohingya people in the country of Myanmar are suffering at this moment caused by the propagation of hate speeches.

In an effort to accuse the Muslim people of the Rohingya in Myanmar of being violent, a series of deceptive images were used to justify the persecution of one of the most segregated ethnic groups in the world.

In 2017, the Reality Check team of the BBC verified that photos and videos that had been recorded decades ago were being used to defame the Rohingya, who have even been denied citizenship in Myanmar.

The extreme violence unleashed from August by the military, with the collaboration of Buddhist crowd agitated by xenophobic monks via Facebook, is the culmination of a systematic process of persecution of this Muslim minority that began in the seventies.

Freedom of expression, mobile phones and this social network are some of the novelties brought about by the democratization of this country with a Buddhist majority. "I fear that Facebook has become a beast, which was not the original intention," Myanmar UN researcher Yanghee Lee said in March in Geneva. Experts in the country agree that Facebook has been the great transmission belt of hatred in Myanmar.

There Internet equals Facebook. The Burmese do not search in the web, they do not go to Google but to Facebook, which even the institutions use (The army uses the facebook post as official way to communicate).

After Mark Zuckerberg boasted in an interview, in the wake of the data leak scandal, of the supposed effectiveness of his company to detect and eliminate hate speech in the Asian country, a group of representatives of civil society Burmese and entrepreneurs published an open letter accusing him of having ignored the warnings they sent him.