The Evolution of Fake News and the Abuse of Emerging Technologies

Roberto Adriani

Abstract

Fake news and post-factual society are quite popular terms today. The literature is investigating this phenomenon from different perspectives. We also know the psychological dimension at the basis of fake news (Lynch M. 2016) and the debate around the need for a new media policy (Goodman E. 2017). However, something else is very important: the evolution process of fake news. Far from being a still life, fake news will evolve and this needs to be monitored closely. In a post-factual society fake news could be fuelled by the abuse of new powerful technologies (Murgia M. and Kuchler H. 2017).

Keywords: fake news; post-truth, AI

1. Introduction

Today, fake news makers have powerful new technologies, such *voice and face morphing*, which make it almost impossible to distinguish truth from falsehood. These technologies will shortly allow the creation of fake videos, in which software makers can superimpose real footage onto a fake audio, using the original voice. The result is footage in which a real person says something he never said. Also, AI may have a dark side and being used by fake news makers *(Vincent J. 2017).*

2. Methodology

The paper, through a narrative approach, investigates these emerging technologies and how they may be misused. The paper analyses scientific articles from international literature, in English, over last ten years. We focus the research on a relatively recent period as fake news is basically a recent phenomenon. Or, to say better, the destabilizing effect of fake news is relatively recent.

The review also covers journalistic articles, which report data, insights or simple news, regarding the two subjects of the key questions. In this case, the paper includes only articles coming from mainstream publications, printed as well as online.

Trying to pursue this objective, articles not clearly reporting the name of the publication, author and date, have been excluded. In addition to that, all the articles have been checked through a web engine search, making sure they have been cited or linked by other mainstream media.

3. Fake news spreading. The lesson we have learnt

Among the many lessons we have learnt about fake news, here are two pillars this paper wants start from.

3.1 Perception and false representation

Fake news is not about simply the perception of reality. Fake news is an objectively false representation of reality and it can be even defamatory. The greatest examples we have seen are from politics, where lies are not merely occasional mishaps, but the core of an electoral strategy.

The other important lesson we have learnt is why people spread fake news. The psychological mechanism behind this, can be reassumed as follows *(Gathman C. 2014).*

a. People do not actually read the content they are sharing

b. People do not consider the legitimacy of specific news sources

c. People are vulnerable to confirmation bias

d. People infer legitimacy from "related content"

e. People see a piece of content as more legitimate the more they see of it

f. People confuse satire and hoax

How can we stop fake news spreading without jeopardising free speech? Who must be in charge of deciding what is fake and what is real? Some experiments are already on the table. Let's see how really promising they are.

United Kingdom

In Great Britain the Brexit referendum has made institutions much more aware of the threat of fake news. Emma Goodman, from the Media Policy Project Blog of the LSE, explains how British institutions try to fake news. An inquiry has been launched by the House of Commons, to investigate this phenomenon, described as a threat to democracy able to undermine the confidence in the media in general. The inquiry has been conducted by the Culture, Media and Sports Committee. Its chair Damian Collins, said that any possible solution must be focused on social media platforms. According to the chair, Facebook is the main social media platform put in the spotlight, and it should put more attention in assessing and notifying fake news to help users. In the end, says consumers should be empowered to assess fake news *(Goodman E. 2017)*.

Germany

It is one of the first laws issued to face fake news, which is why, as Emma Thomasson from Reuters explains, it can be considered an international experiment. The law, which came into full force on Jan. 1, 2018, aims to ensure Germany's tough prohibitions against hate speech, including pro-Nazi ideology, are enforced online by requiring sites to remove banned content within 24 hours or face fines of up to 50 million euros ($62 million). The law, called NetzDG for short, is an international test case and how it plays out is being closely watched by other countries considering similar measures. In addition to that, social media companies are asked to do more to stop fake news. The lawmakers are also pushing for social media firms to set up an independent body to review and respond to reports of offensive content from the public, rather than the individual companies doing that themselves. The law is now to be amended, following the criticism from opponents of the law, including free speech campaigners and the Association of German Journalists, who say the threat of hefty fines is prompting internet firms to err on the side of caution and block more content than is necessary. German authorities have stressed, however, that they believe the law is working well in terms of forcing social media companies to delete offensive posts *(Thomasson E. 2018,* https://www.reuters.com/article/us-germany-hatespeech/germany-looks-to-revise-social-media- law-as-europe-watches-idUSKCN1GK1BN*)*.

Italy

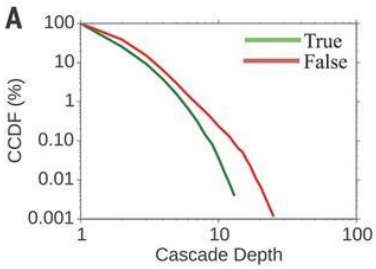
Here the effort is mainly focused on education and training of the younger generation. The former presidency of the Lower House, together with the Ministry of Education, launched an extraordinary experiment in cooperation with leading digital companies including Facebook. The effort was focused on training a generation of students steeped in social media how to recognize fake news and conspiracy theories online. The program tried to deputize students as fake-news hunters, showing them how to create their own blogs or social accounts to expose false stories and 'showing how you uncovered it *(Horowitz J. 2017,* https://www.nytimes.com/2017/10/18/world/europe/italy-fake-news.html*)*. In addition to that, Facebook rolled out for its Italian users a new fact-checking program aimed at identifying and debunking false information that appears on the site. Like similar efforts Facebook has launched in the past, the program relies on user reporting and third-party fact checkers to flag potential false material *(Serhan Y. 2018* https://www.theatlantic.com/international/archive/2018/02/europe-fake-news/551972*/)*. In any case, this programme does not seem having achieved great results.

3.2 Fake news spreads faster than truth

Another important lesson is about how fast fake news spreads, even more than truth. It is the case to briefly report here a recent study on that, published by Science, on March 2018. It analyses fake news spread via Twitter. The author's assumption is that a rumor cascade begins on Twitter when a user makes an assertion about a topic in a tweet, which could include written text, photos, or links to articles online. Others then propagate the rumor by retweeting it. A rumor's diffusion process can be characterized as having one or more cascades, defined by authors as instances of a rumor spreading pattern that exhibit an unbroken retweet chain with a common, singular origin. So, if a rumor "A" is tweeted by

10 people separately, but not retweeted, it would have 10 cascades each of size one. Conversely, if a second rumor "B" is independently tweeted by two people and each of those two tweets is retweeted 100 times, the rumor would consist of two cascades, each of size 100.

When the authors analysed "the diffusion dynamics of true and false rumors, they found that falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information. A significantly greater fraction of false cascades than true cascades exceeded a depth of 10, and the top 0.01% of false cascades diffused eight hops deeper into the Twittersphere than the truth, diffusing to depths greater than 19 hops from the origin tweet (Fig. 1, 2A in the original).

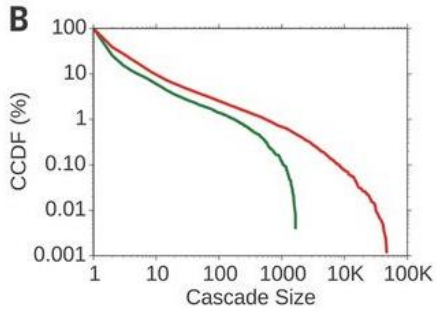


(Fig. 1, 2A in the original)

Falsehood also reached far more people than the truth. Whereas the truth rarely diffused to more than 1000 people, the top 1% of false-news cascades routinely diffused to between 1000 and 100,000 people (Fig 2B in the original).

Falsehood reached more people at every depth of a cascade than the truth, meaning that more people retweeted falsehood than they did the truth. The spread of falsehood was aided by its virality, meaning the falsehood did not simply spread through broadcast dynamics but rather through peer-to-peer diffusion characterized by a viral branching process.

It took the truth about six times as long as falsehood to reach 1500 people and 20 times as long as falsehood to reach a cascade depth of 10".



(Fig. 2, 2A in the original)

Another important result highlighted by this study, is that "contrary to conventional wisdom, robots accelerated the spread of true and false news at the same rate, implying that humans, not robots, are more likely responsible for the dramatic spread of false news". The authors also checked possible bias.

"In case there was concern that the authors' conclusions about human judgement were biased by the presence of bots in their analysis, they employed a sophisticated bot-detection algorithm, to identify and remove all bots before running the analysis. When they added bot traffic back into the analysis, they found that none of their conclusions changed. False news still spread farther, faster, deeper and more broadly than the truth in all categories of information.

Although the inclusion of bots accelerated the spread of both true and false news, it affected their spread roughly equally *(Vosoughi S. et al., 2018, p. 2,3,5)*.

4. The abuse of emerging technologies

Some new technologies can push us into a new era of fake news, in which manipulation looks ever more believable.

4.1 Do not believe everything you see

Unless I see the nail marks in his hands and put my finger where the nails were, and put my hand into his side, I will not believe it, said Saint Thomas It is a famous sentence, which highlights how important is for us to see (and even to touch) to believe.

Unfortunately, audio and video manipulation is questioning this assumption. As Olivia Solon reports in The Guardian. There is a new breed of video and audio manipulation tools, made possible by advances in artificial intelligence and computer graphics, that will allow for the creation of realistic looking footage of public figures appearing to say, well, anything. Nothing is sure from now on. We have long been told not to believe everything we read, but soon we will have to question everything we see and hear as well. An important experiment has been set up at Stanford University, where a software "is able to manipulate video footage of public figures to allow a second person to put words in their mouth – in real time. Face2Face captures the second person's facial expressions as they talk into a webcam and then morphs those movements directly onto the face of the person in the original video. The research team demonstrated their technology by puppeteering videos of George W Bush, Vladimir Putin and Donald Trump. On its own, Face2Face is a fun plaything for creating memes and entertaining late night talk show hosts. However, with the addition of a synthesized voice, it becomes more convincing – not only does the digital puppet look like the politician, but it can also sound like the politician. Similar research has been conducted at the University of Alabama. With 3-5 minutes of audio of a victim's voice – taken live or from YouTube videos or radio shows – an attacker can create a synthesized voice that can fool both humans and voice biometric security systems used by some banks and smartphones. The attacker can then talk into a microphone and the software will convert it so that the words sound like they are being spoken by the victim – whether that's over the phone or on a radio show. Not only universities are interested in developing these technologies. Canadian startup Lyrebird has developed similar capabilities, which it says can be used to turn text into on-the-spot audiobooks 'read' by famous voices or for characters in video games.

But the most impressive, and cited, experiment, is about Obama. The University of Washington's Synthesizing Obama project (Fig. 3), where they took the audio from one of Obama's speeches and used it to animate his face in an entirely different video with incredible accuracy (thanks to training a recurrent neural network with hours of footage), to get a sense of how insidious these adulterations can be *(Solon O., 2017* https://www.theguardian.com/technology/2017/jul/26/fake-news-obama-video-trump-face2face-doctored-content *).*

(Fig. 3)

available at https://www.theguardian.com/technology/2017/jul/26/fake-news-obama-video- trump-face2face-doctored-content

But the problem is not just the proliferation of falsehoods, as reported by Franklin Foer in The Atlantic. Fabricated videos will create new and understandable suspicions about everything we watch. Politicians and publicists will exploit those doubts. When captured in a moment of wrongdoing, a culprit will simply declare the visual evidence a malicious concoction.

The article points out another risk. In other words, manipulated video will ultimately destroy faith in our strongest remaining tether to the idea of common reality. A sort of new LSD, as the author says? Fake-but-realistic video clips are not the end point of the flight from reality that technologists would have us take. The apotheosis of this vision is virtual reality. VR's fundamental purpose is to create a comprehensive illusion of being in another place. With its goggles and gloves, it sets out to trick our senses and subvert our perceptions. Video games began the process of transporting players into an alternate world, injecting them into another narrative. But while games can be quite addictive, they aren't yet fully immersive. VR has the potential to more completely transport—we will see what our avatars see and feel what they feel.

Maybe we will find a way to cope with that but, in the meantime, it would be better to be more prudent. Perhaps society will find ways to cope with these changes. Maybe we will learn the scepticism required to navigate them. Governments have been slow to respond to the social challenges that new technologies create, and might rather avoid this one. The question of deciding what constitutes reality isn't just epistemological; it is political and would involve declaring certain deeply held beliefs specious *(Foer F., 2018* https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/*)*.

4.2 The dark side of AI

The response Hillary Clinton got when her book debuted on on Amazon's Web was surprising. Of the 1,600 reviews posted on the book's Amazon page in just a few hours, the company soon deleted 900 it suspected of being bogus: written by people who said they loved or hated the book, but had neither purchased nor likely even read it. Fake product reviews—prompted by payola or more nefarious motives—are nothing new, but they are set to become a bigger problem as tricksters find new ways of automating online misinformation campaigns launched to sway public opinion.

At the University of Chicago some researchers are investigating whether artificial intelligence could be used to automatically crank out bulk reviews that are convincing enough to be effective. Their latest experiment involved developing AI-based methods to generate phony Yelp restaurant evaluations. (Yelp is a popular crowdsourced Web site that has posted more than 135 million reviews covering about 2.8 million businesses since launching in July 2004). The researchers used a machine-learning technique known as deep learning to analyze letter and word patterns used in millions of existing Yelp reviews.

Deep learning requires an enormous amount of computation and entails feeding vast data sets into large networks of simulated artificial "neurons" based loosely on the neural structure of the human brain. The Chicago team's artificial neural network generated its own restaurant critiques—some with sophisticated word usage patterns that made for realistic appraisals and others that would seem easy to spot, thanks to repeated words and phrases.

But when the researchers tested their AI-generated reviews, they found that Yelp's filtering software—which also relies on machine-learning algorithms—had difficulty spotting many of the fakes. Human test subjects asked to evaluate authentic and automated appraisals were unable to distinguish between the two. When asked to rate whether a particular review was 'useful', the human's respondents replied in the affirmative to AI-generated versions nearly as often as real ones *(Greenemeier L., 2017* https://www.scientificamerican.com/article/could-ai-be-the-future-of-fake-news-and-product-reviews/*)*.

In a nutshell, the authors showed how a two phased review generation and customization attack can produce reviews that are indistinguishable by state-of-the-art statistical detectors. They conducted a survey-based user study to show these reviews not only evade human detection, but also score high on 'usefulness' metrics by users.

The authors also say that AI can not only assist fake news detection but also generate fake news. Given the availability of large-scale news datasets, an attacker can potentially generate realistic looking news articles using a deep-learning approach (RNN). And due to its low economic cost, the attacker can pollute social media newsfeeds with a large number of fake articles.

The researchers end hoping their results "will bring more attention to the problem of malicious attacks based on deep learning language models, particularly in the context of fake content on online services, and encourage the exploration and development of new defences (*Zhao B.Y et al., 2018, P. 1,13)*.

It is an additional warning about how fake news may become a dirty tool, used by dishonest companies to strike at their competitors' reputation. Lil Miquela is another example about how new technologies, AI in particular, can manipulate the social media sphera.

Lil Miquela has been a source of fascination for many on Instagram since not long after her account launched in April 2016, but for her first two years of existence, no one could definitively say who or what was behind the operation. The Bermuda hack-slash-PR-stunt solved at least part of the mystery, linking Miquela to Brud, a Los Angeles-based startup that specializes in "robotics, artificial intelligence and their applications to media businesses"—but the entire saga remains a master class in postmodern performance art. The author says that Lili Miquela Instagram profile is potentially money-making. Miquela isn't just a flashy stunt: She has serious money-making potential. Already, the virtual influencer has partnered with Giphy and Prada and posed wearing Diesel and Moncler. In February, Miquela said she had never been paid to model a piece of fashion on her feed, but that could change at any moment. (Lil Miquela's PR representatives did not respond to queries about whether she has posted any sponsored content since that statement).

The appearance of influencers has generated new questions about how to distinguish advertising from paid social media influencers. The story of Lil Miquela is even a step forward: what about if the influencer is a non-existing person, but only a virtual profile AI based? But virtual models and influencers like Lil Miquela raise thorny questions. Last year, the Federal Trade Commission updated its endorsement guides to require influencers to disclose their marketing relationships and identify paid posts with a hashtag like #ad or #sponsored—but it's not clear how those rules would apply to influencers who aren't human, and whose backers, like Lil Miquela's, are shrouding themselves in mystery *(Katz M., 2918,* https://www.wired.com/story/lil- miquela-digital-humans/*)*.

The final question, in this case, is about why we should trust the opinion of someone who does not exist?

(Fig. 4)

Lil Miquela's profile, from the Instagram app of the author.

5. Key findings

The key findings can be reassumed into two points

5.1 Fake news still a problem

This phenomenon is still dangerous, in spite of all efforts. On the other hand, this does not mean we have to give up. A solution is currently being sought, both from a technological (can block-chain technology be used to stop fake news?) as well as from a contents point of view.

In any case, education, training and digital literacy of children must be part of the solution.

5.2 New threats

The new threats are represented by a dishonest (or even criminal) use of AI, which allow fake news makers to create more pervasive and dangerous hoaxes. Unlike "traditional" fake news, text based, the new ones are based on the power of the image (CGI). It is particularly insidious as it is grounded on the general belief that everything you can see is true.

References

[1] Lynch, M. P. (2016), "The Internet of Us: Knowing More and Understanding Less in the Age of Big Data", ISBN 9780871406613 Liveright, New York, NY.
[2] Goodman E. (2017), "How has media policy responded to fake news?", London School of Economics-Media Policy Project Blog, available at: http://blogs.lse.ac.uk/mediapolicyproject/2017/02/07/how-has-media-policy-responded-to-fake- news/
[3] Murgia M. and Kuchler H. (2017), "Facebook struggles to purge fake news," Financial Times, 1 May 2017, available at: https://www.ft.com/content/0feeafe6-2c01-11e7-9ec8-168383da43b7
[4] Vincent J. (2017), "New AI research makes it easier to create fake footage of someone speaking", The Verge, 12 July 2017, available at: https://www.theverge.com/2017/7/12/15957844/ai-fake- video-audio-speech-obama
[5] Gathman C. (2014), "Why people fall for dumb Internet hoaxes", The Washington Post, 12 September 2014, available at: https://www.washingtonpost.com/news/the- intersect/wp/2014/09/12/why-people-fall-for-dumb-internet- hoaxes/?noredirect=on&utm_term=.daae5b332eb3
[6] Thomasson E. (2018), "Germany looks to revise social media law as Europe Watches", Reuters, 8 March 2018, available at: https://www.reuters.com/article/us-germany-hatespeech/germany- looks-to-revise-social-media-law-as-europe-watches-idUSKCN1GK1BN
[7] Horowitz J. (2017), "In Italian Schools, Reading, Writing and Recognizng Fake News", The New York Times, 18 October 2017, available at: https://www.nytimes.com/2017/10/18/world/europe/italy- fake-news.html
[8] Serhan Y. (2018), "Italy scrambles to fight misinformation ahead of its elections" The Atlantic, 24 February 2018, available at: https://www.theatlantic.com/international/archive/2018/02/europe- fake-news/551972/
[9] Vosoughi S. et al. (2018), "The spread of true and false news online", Science, 09 Mar 2018: Vol. 359, Issue 6380, pp. 1146-1151, DOI: 10.1126/science.aap9559
[10] Vosoughi S. et al. (2018), "The spread of true and false news online", MIT Initiative on the Digital Economy, 8th March 2018, pages. 1-5.
[11] Solon O. (2017), "The future of fake news: don't believe everything you read, see or hear", The Guardian, 26 July 2017, available at https://www.theguardian.com/technology/2017/jul/26/fake- news-obama-video-trump-face2face-doctored-content
[12] Foer F. (2018), "The Era of Fake Video Begins. The digital manipulation of video may make the current era of fake news seem quaint", The Atlantic, May 2018, available at https://www.theatlantic.com/magazine/archive/2018/05/realitys-end/556877/
[13] Greenemeier L. (2017) "Could AI Be the Future of Fake New and Product Reviews?", Scientific American, 16 October, 2017, available at https://www.scientificamerican.com/article/could-ai-be- the-future-of-fake-news-and-product-reviews/
[14] Zhao B.Y et al. (2018), "Automated Crowdturfing Attacks and Defenses in Online Review Systems", University of Chicago.
[15] Katz M. (2018), "CGI 'influencers like Lil Miquela are about to flood your feeds. The full truth behind Lili Miquela's account may never become clear-but when it comes to confusing encounters with hyper-realistic digital humans, she's only the beginning", Wired, 1 May, 2018, available at https://www.wired.com/story/lil-miquela-digital-humans/