

COMBATING MISINFORMATION & DANGEROUS SPEECH



UniMAC
G13

Written by:
Nordor Harriet
(DRIMF-GIJ)



Combating Misinformation and Dangerous Speech

July 2023

Published by
Paradigm Initiative

Author

Nodor Harriet

Design and Layout

Kenneth Oyeniya, Communications Officer, Paradigm Initiative



Creative Commons
Attribution 4.0 International (CC BY 4.0)

EXECUTIVE SUMMARY

In this fast-rising era of technology, social media has fundamentally changed the way people communicate and receive information around the world. The same platform that gives access to receive and share information has also been a key to threatening, harassing, and manipulating people with false content via speech, texts, images, or videos that increase the rate of violence within targeted groups and cause depression to victims. The purpose of this research is to create awareness of misinformation and dangerous speech in the digital age and its effects. The introduction focuses on the history of misinformation and dangerous speech with the types and modes of spreading in the digital space. Further analysis follows on two of social media's communication giants, Facebook and Twitter as fake news sites will be spotted. A further look will be at Ghana's digital space and relevant examples of misinformation and dangerous speech highlighted. Finally, the effects of misinformation and dangerous speech on victims will be discussed, and some Ghanaian laws, such as the Electronic Communications Act, 2008(Act 775), that govern misinformation and dangerous speech will be analysed with solutions to solve issues relating to misinformation and dangerous speech.

BACKGROUND

Information is consumed daily on social media platforms by millions of citizens. It changes the media landscape in relation to the content shared on television, radio, and print. Unfortunately, the digital age, which has brought some form of relief and joy to the citizenry, has also been the same ground on which false news, misinformation and content are easily created and spread to people breeding hatred, mistrust, and chaos in some parts of the country. Misinformation is incorrect or misleading information that is deliberately deceptive. Rumours are information not attributed to any source and so unreliable and often unverified but can turn out to be either true or false.¹

Also, misinformation means giving wrong information, which can mislead the receiving audience. With regards to this, news or analysis that misinforms tends to malign and denigrate the image and integrity of the targeted individual or group, making the affected person or persons lose their dignity. The history of misinformation can be dated back to the year 1835 when the New York Sun Publication claimed that there was an alien civilization on the moon.²

In recent times, due to the advancement of technology and digitization, misinformation has grown stronger with the aim of spreading fake news, dangerous speech, and propaganda. Fake news can be described as stories that are published and are not fact-based. The aim of posting such content is to mainly make money after people click on links because of titles that are appealing (clickbait). Clickbait refers to content whose main purpose is to attract attention and encourage visitors to click on a link on a particular webpage.³ In some places like the United States of America, the practice is considered unethical as it exposes celebrities and politicians to unproven allegations.¹

TYPES OF MISINFORMATION

Different kinds of misinformation vary significantly in their tactics, intent and impact. The following are five different types of misinformation.

1. Satire
2. Fabricated Content
3. Manipulated Content
4. Imposter Content
5. False Content⁴

News Satire is a form of content that typically makes fun of news programs and uses humor to engage with their audience⁵. The goal is to provide entertainment and suggest humorous critiques of political or pop cultural events. Sadly, a satire which drives from a harmless point of view has undergone a metamorphosis into parody presented in the format that describes a piece of news or information in which human or individual vices, follies, abuses, or shortcomings are held up to censure by means of parody, irony, ridicule, derision, burlesque, caricature, or other methods to make up an entirely fictitious content.⁶

While the above play on humorous attacks, fabricated content is a type of information with no factual basis; however, the stories are published to create legitimacy and are often believed to

1 <https://en.wikipedia.org>

2. A Brief History of Fake News; www.cits.ucbs.edu

3. Oxford Online Dictionary

be a trustworthy source because partisan organizations often present information with some neutrality⁷. The aim of such content is for the purpose of financial gain through artificially intelligent online sources. It is important to realize that it is not just the public who can be misled by false information. Unfortunately, credible news outlets are sometimes misled and may not realize it until it is too late. While fabrication is 100% false, manipulation has some truth to it. Manipulation can be described as news stories that use real images or videos to create a false narrative⁸. Despite there being some truth, using an adaptation of imagery to sensationalize a story still misleads consumers by developing a false connection. Imposter content of misinformation uses a well-known name, brand or logo to lure people into believing that whatever newsfeed is shared on social media handles and portals is the truth, and this occurs by using images, videos, quotes, and speeches, among others, to change the meaning of the message and spread confusion.

DANGEROUS SPEECH

Dangerous speech is any form of expression (speech, text, or image) that can increase the risk that its audience will condone or commit violence against members of another group.⁹In the early 2000s, Benesch noticed striking similarities in the rhetoric that political leaders in many countries used, during the months and years before major violence broke out. In the definition of dangerous speech, violence means direct physical (or bodily) harm inflicted on people, not other forms of harm such as doxing, incitement to self-harm, discrimination, or social exclusion. These other forms of harm are important, and of course, dangerous speech may inspire people to inflict many forms of harm.

DANGEROUS SPEECH ONLINE –THE ROLE OF SOCIAL MEDIA

As stated earlier in the definition, dangerous speech commits and condones violence in the sense that in the most large-scale violence between people, only a small proportion (usually young men) carry out violence and people close to them (siblings, teachers, and friends) often condone or even encourage it. Generally, when a society suffers major intergroup violence, a few commit it and a much larger number condone it.

Digital media and the internet have immeasurably changed the way people spread all kinds of messages, from the innocuous to the incendiary. Those who seek to turn groups of people violently against each other can spread Dangerous Speech quickly – especially in places where there is already a risk of mass violence. Ideas and narratives once confined to the fringes of popular discourse – including extremist ideas are now widely available. Speakers who could hardly find an audience offline, even those who espouse the most widely derided ideologies, can find at least a few fellow-thinkers across the world, and can form so-called “echo chambers” in which they bolster and further radicalize each other. By forging such bonds, people can collectively disseminate harmful content further than they could have alone and with the fervour of solidarity. Others are motivated neither by hatred nor conviction, but by simply wanting more followers and/or more money (from subscribers or advertisers).

Online, people can also communicate anonymously. On social media platforms like Twitter or Reddit, or messaging like platforms like WhatsApp or Discord, they can spread ideas that they might not dare to express offline, where their identities would be known.

As it has become increasingly obvious that online content leads to serious offline harm, governments, researchers, activists, and internet companies have sought ways to diminish the problem. The first, most obvious response is simply to remove bad content or censor it. Each country has laws

prohibiting certain forms of speech (they vary) and social media companies like Facebook and Twitter also have their own rules forbidding certain kinds of content, such as hate speech, nudity, or incitement to violence¹⁰.

Censorship, whether by governments or private companies, poses significant risks to democracy and freedom of expression since it is almost impossible to do it without making serious mistakes. First, although some content is obviously harmful or even illegal, most is quite context-dependent or ambiguous, and it is often difficult to agree on where to draw the lines. Secondly, policing the internet for harmful content is a job so huge that its scale is hard even to imagine: everyday, 147 billion people log into Facebook alone and post billions of pieces of information¹¹. Although internet companies train thousands of people (often ill-paid and psychologically battered from looking at terrible content all day) to decide which posts to take down at such a scale, mistakes are inevitable and numerous¹².

RESPONDING TO HATEFUL AND DANGEROUS SPEECH ONLINE.

There are also many other ways of diminishing harmful content or its damaging effects. One might try to persuade people to stop posting such content in the first place (a preventive approach, rather than a reactive one like deletion), or support those who are attacked by it. Internet users themselves (not governments or companies) are conducting many ingenious experiments in responding to harmful content online. A single response to harmful speech online – deletion, or takedown is practically by far the most discussed and demanded, but other responses deserve notice, especially those that convince people to post less harmful speech. Takedown focuses on the offending content alone – not on those who post it nor those who are harmed by it – so it does not do much to persuade people to stop thinking, speaking, or reposting in harmful ways, and it does not remedy the harm done to people who are exposed to the content before it is removed. Takedown cannot keep up with the staggering rate at which new content appears online, except perhaps if takedown becomes automated and algorithmic prior to censorship, which tends to be overboard and would infringe on freedom of speech. Also, takedown is a method that can be practised only by internet companies making and applying their own internal rules and polices, while governments pressure them to take down more content.

Blockbots are one more tool that the targets of harassment use to protect themselves against Dangerous Speech, in this case by preventing themselves from seeing it at all. Blockbots are applications that allow Twitter users to block lists of other users – and to share those lists easily. When a user subscribes to a list, the bot uses the Twitter API to block the accounts on the list automatically, avoiding the tedious process of blocking each account individually. This innovation proved so useful for targets of harassment that Twitter built into its platform the option to import and export block lists in 2015.¹³

These efforts use fundamentally different methods to diminish harm: from magnifying racist messages and displaying them, literally in the sunlight to try to leave them hidden in the dark corners of Twitter. Each of them deserves study, to determine its actual capacity to reduce harmful speech or the damage it does.

ANALYZING THE HATE AND COUNTER SPEECH ACCOUNTS ON TWITTER

The online proliferation of hate speech has caused several countries and companies to implement laws against hate speech to enforce citizens to restrain from such behavior. Countries such as Germany, the United States of America, France, and others have laws banning hate speech. Social

media sites such as Facebook and Twitter usually respond to hate speech with the suspension or deletion of the message or the user account itself. While these laws may reduce the amount of hate speech in online media, it does so at the cost of causing harm to the freedom of speech. Another potential alternative to tackle hate speech is “counter speech”. The main idea behind counter-speech is to add more speech to the conversation and try to change the mindset of the hate speaker. There are several initiatives with the aim of using counter-speech to tackle hate speech. For example, the Council of Europe supports an initiative called the ‘No Hate Speech Movement’ with the aim to reduce the levels of acceptance of hate speech to develop online youth participation and citizenship, including in Internet governance processes. UNESCO released a study (Gagliardone et al, 2015) titled ‘Countering Online Hate Speech’ to help countries deal with this problem. Social platforms like Facebook have started counter-speech programs to tackle hate speech. Facebook has even publicly stated that it believes counter-speech is not only potentially more effective, but also more likely to succeed in the long run.¹⁴ Combating hate speech in this way has some advantages: it is faster, more flexible, and responsive, capable of dealing with extremism from anywhere and in any language and it does not form a barrier against the principle of free and open public space for debate.¹⁵

Observations from this analysis show that hateful accounts tend to express more negative sentiments and profanity in general. If the hateful tweet is from a verified account, it seems to be much more viral as compared to other hateful tweets. A perfect example in the Ghanaian context is the issue of a young Ghanaian blogger named Albert Nat Hyde popularly referred to as Bongo Ideas who used his Twitter handle to spread falsehood against three renowned female media personalities namely Nana Aba Anamoah, Serwaa Amihere and Bridget Otoo with actress Lydia Forson who was not left out as he threw shades of hate speech about them peddling falsehood about their lifestyle and addressing them as disappointed feminists. Tweets from Albert’s handle showed derogatory comments about these four women questioning their integrity in a hateful manner and shading them into a dark light resulting in cyberbullying. In a viral video, the young gentleman was seen being interviewed for a reality TV show of which the three media personalities were judges. Nana Aba Anamoah, one of the victims of the cyberbullying messages exposed his Twitter account and reading his tweets saw falsehood peddled in the tweets. In questioning him as to how and where he gets his information to malign and bully their personalities online, Albert replies by telling them that Twitter is a different space from the real world.

Recently, Bongo Ideas wrote on his Twitter page that Bridget Otoo got married to cover up her pregnancy as rumours went around. Bridget got married in a colourful ceremony held in Takoradi and was graced by some friends, family and well-wishers. Albert’s tweet stated “after saying shitty things about men in the name of feminism, she has begged for one. He then added “When their values start to depreciate with time, they start searching for a man to share their toxicity with. Sorry for you, man”. Yvonne Nelson, a Ghanaian actress took to her Twitter handle a message telling critics to just ignore her tweets if they do not agree with her after she posted a tweet of a gentleman who dropped a hate speech comment about her and wanted to be her personal assistant. After this incident, a lot of young people found themselves deleting tweets that contained hate speech targeting race and physical traits such as obesity, the LGBTQI+ community, nationality and religion. It soon became a ground for change as Albert was offered the chance to spearhead an

¹³ Perspectives on Harmful Speech Online: A Collection of essays pdf, 2017 www.newslibraryproject.com.

¹⁴ Bartlett and Krasodowski-Jones, 2015

¹⁵ Analyzing Hate and Counter Speeches on Twitter pdf by Matthew et al.

¹⁶ de Lima et al, 2018

¹⁷ (@Bongo Ideas) Twitter handle of Nat Hyde

anti-cyberbullying campaign which he declined due to personal reasons.

In terms of personality traits, the counter speakers seem to have a higher quotient of 'agreeableness'. They are more altruistic, modest, and sympathetic. The hateful users, on the other hand, seem to have a higher quotient of 'extraversion' indicating that they are more energetic and talkative in nature. They are more cheerful, excitement-seeking, outgoing and sociable in nature.

The harsh nature of Twitter on hate speech has led to several user account suspensions and deletions. This is true for other social media sites as well. Due to this, several new social media sites have sprung up in recent years like Gab's Wrong think which support free speech on their site and allow users to post content that would not be allowed on other sites. This has resulted in gab becoming an echo chamber for right-leaning dissemination.

DETECTING FAKE NEWS ON FACEBOOK.

The dissemination of misinformation has always been a feature of society. However, the ubiquity and potential damage that such misinformation or 'fake news' can have is elevated by the emergence of social media platforms such as Facebook, which engages a global audience. Fake news can take several forms ranging from instances in which visual or textual information is inserted into an article in order to subtly bias an argument one way or the other, to wholly fabricated content, often including extraordinary claims created to and shared systematically with the sole intention to deceive. Fake content captures attention and spreads rapidly in comparison to real news and such items often employ emotionally charged language. Research suggests that such emotional content could be one of the key factors which prevent social media users from engaging in a critical assessment of the core message. However, one psychological concept known as 'emotional intelligence (EQ)', could attenuate this effect.

One of the engaging aspects of social media is the rapidity with which news and information is updated, news posts from information content providers compete with each other to create highly salient 'bite-sized' headlines which can be rapidly digested before moving to the next and this format is well suited to users who wish to be presented with fresh content which reflects their existing attitudes, fears and implicit prejudices and capitalizes on their existing emotional connections to particular topics (for example, political affiliation). In contrast, this type of news platform is ill-suited to the type of critical thinking and analytical reasoning that is required to judge the truthfulness of content.

¹⁷ Detecting Fake News on Facebook: The role of Emotional Intelligence pdf by Preston et al, <https://doi.org/10.3771/journal.pone.0246757>

¹⁸ Electronics Communication Act, 2018 (Act 775 section 76)

How To Combat Misinformation and Dangerous Speech Online

With all factors drawn and suggested to detect misinformation and dangerous speech, netizens have the power to curb and if possible, avoid misinformation by fact-checking news stories from outlets and social media platforms by paying attention to the kind of headline stories, grammatical constructions and validity with a bias-free format. Section 76 of the Electronics Communications Acts, 2008, Acts 775 of Ghana states as follows;

“Knowingly sending a communication which is false or misleading and likely to prejudice the efficiency of life saving service or to endanger the safety of any person, ship, aircraft, vessel or vehicle by means of electronic communication commits an offense and is liable on summary conviction to a fine of not more than three thousand penalty units or to a term of imprisonment of not more than five years or both.¹⁸

This best explains how perpetrators of spreading fake news face the law and even have their websites even closed. In the same law, Sections 74 and 75 deal with the intentional spread of false news and signals that cause fear and panic with charges of a fine penalty of not more than one thousand penalty units or a term of imprisonment of not more than three years or both and a fine penalty of not more than three thousand units or a term of imprisonment of not more than five years or both respectively. However, these provisions have a chilling effect on media freedoms as civil sanctions must be preferred. On the part of the media outlets, intensive awareness must be created to inform and educate citizens on the dangers and penalties associated with spreading misinformation and dangerous speech and how to avoid clickbait links that spread violence.

Media outlets must exhibit fairness in news findings, content creation, promotions, and advertisements among others without the expression of sentiments and attached exaggerations. Also, having a robust accreditation system is needed as the accreditation board creates a system of codes of conduct to put workers who fall below the standards under check and penalize them according to these laws. Media or journalism organizations should also build accreditation boards to check media outlets that create a system of libel, plagiarism, cyber-bullying and other suits of malpractices.

CONCLUSION

Although the media sometimes operates under the an environment of political power (be it government or opposition powers), netizens should be alert to the kinds of sites and information they feed on to avoid violence and chaos. Schools should make it a habit of introducing digital rights courses in their IT sessions to equip children to prevent or avoid the menace. Counsellors and psychologists should be trained in digital rights to aid victims suffering from depression and low self-esteem to recover from the effects of misinformation and dangerous speech. Intense Awareness Creation must be enforced by Municipal and district executives in towns, villages and cities to inform the community of the causes and effects of misinformation and dangerous speech to reduce all forms of trauma. Fact-checking must be encouraged across all sectors in society. To make the digital space a haven for all, sites that propagate falsehoods must be reported to the Cyber Security Department of the Police Service.



© 2023 Paradigm Initiative
HQ: 374 Borno Way, Yaba, Lagos - Nigeria.