# Digital Technologies Used to Combat Disinformation and Fake News

**Andreea Florina RADU[1], Ioana PETCU[2]**
[1]"Babes-Bolyai" University of Cluj-Napoca
[2]National Institute for Research and Development in Informatics - ICI Bucharest
andre.radu@gmail.com, ioana.petcu@ici.ro

**Abstract:** One fundamental principle of free democratic societies that is protected by constitutions is the freedom of expression. To form opinions on a range of economic, social, and political issues, citizens in healthy democratic societies must be able to access a large amount of information, including that found online. However, given the increasing prevalence of fake news it is crucial that the information that citizens receive through various channels – such as social media posts, online news articles, and platforms – is verifiable. The dissemination of false information can have a variety of negative effects, including undermining democracy, polarizing discussions, and posing risks to people's health, safety, and environment. While propaganda has taken many forms over time, digital "fake news" is growing increasingly dangerous. This article raises the alarm about the extent of the danger to which users of the online space are exposed in terms of fake news and the digital means by which this information is disseminated, and outlines some solutions to combat it by raising public awareness.
**Keywords:** Online space, Fake news, Disinformation, Misinformation, Digital technologies and protocols.

## INTRODUCTION

A story may be posted online by anyone, and social media makes news spread quickly. Therefore, it is imperative that people critically evaluate the information they read and see before placing their trust in its accuracy.

The occurrence of fake news is as ancient as the concept of news itself, however referred to by many terms throughout history, such as rumor, deception, diversion, and propaganda.

The proliferation of false information and deceptive news is a significant challenge due to advancements in technology and the expansion

https://doi.org/10.54851/v6i1y202402

of mass communication methods and data transmission channels in today's interconnected societies. While propaganda has existed for a long time, the rise of digital "fake news" poses a significant danger, mostly because it is now much easier to create, spread, and consume such content. Specifically, this refers to the shift from relying solely on print media to utilizing online platforms, social networks, blogging, vlogging, video sharing, and so on.

It should be noted that there is no universally accepted definition of *fake news* (Lazer, 2018), being rather a broad term, which encompasses different concepts such as propaganda, disinformation, rumors, manipulation and fictitious articles (Aldwairi & Alwahedi, 2018). However, it can be understood as news that is entirely false or contains partially true information presented in a distorted, exaggerated, or out-of-context manner. The purpose of fake news is to manipulate readers and create false beliefs.

*Disinformation*, as defined in UNESCO's handbook for journalism education "Journalism, ‚Fake News' and Disinformation" (Ireton & Posetti, 2018) refers to deliberated and often orchestrated efforts to confuse or manipulate individuals, by providing them with dishonest information. On the other hand, misinformation refers to misleading information that is created or spread without any manipulative or malicious intent.

For a better understanding of *misinformation* and *disinformation*, the explanation given by Staats (2021) can be considered, who differentiates between the two as follows: *fake news* imitates the structure of mainstream news, whereas disinformation does not. Furthermore, *disinformation* is intentionally created to deceive, whereas those involved in *misinformation* are unaware that the information they are sharing is counterfeit. Staats also addresses the term of disinformation which refers to the deliberate spreading of secret information, often by altering its context, date, or time, with the aim of serving personal or corporate interests rather than those of the public. Therefore, fake news and disinformation yield misinformation.

## THE THREATS OF DISINFORMATION AND FALSE INFORMATION

The World Economic Forum has identified fake news as a prominent menace to society because they create and maintain certain anxieties and social tensions, alarm the citizens, inoculate the feeling of distrust in the state institutions and create the premises for several forms of undermining their authority and even sabotage.

Misinformation and disinformation are used by various persons and organizations, including political factions and extremists. They can employ technology to simulate the appearance of their content originating from a reliable source and being supported by verifiable proof. Additionally, they can generate multiple social media profiles to disseminate false information around Internet, resulting in the widespread popularity of the content. By employing predictive algorithms, these perspectives can be disseminated directly to persons who are most inclined to propagate and strengthen those messages. Their dissemination can result in several detrimental outcomes, including endangering democratic systems, exacerbating divisive discussions, and jeopardizing the well-being, safety, and ecological conditions of citizens.

A great challenge is represented by the dimension of visibility that they acquire as a result of the support received through the circulation of messages (their propagation), starting from taking over, distributing, quoting and redistributing.

During times of conflict, for examples, strategies known as grey-zone tactics, such as propaganda and disinformation, are used to gain a strategic edge. Thus, the use of disinformation can have a crucial impact in any conflict, without the need for physical aggression. In fact, more than a century ago, the historian Marc Bloch, in his well-known article in specialized literature, "Reflections d'un historien sur les fausses nouvelles de la guerre" (1921), states that fake news, during the First World War, "disturbed the minds of the masses, sometimes overstimulating them, sometimes breaking their courage, their variety, strangeness

and power continuing to amaze anyone who can remember and remembers that believed them".

In reality it is rather difficult to identify, track and control unreliable content and mostly to deal with its widespread and this challenge asks for collaborative and synchronized efforts of all stakeholders involved, namely governments, online platforms, news media, NGOs, and individuals.

## TOOLS FOR LAUNCHING AND DISTRIBUTING DISINFORMATION AND FAKE NEWS

### Tools with direct human involvement:

**TROLLING** (Social Media Victims Law Center, 2024) is an activity carried out by an individual who deliberately seeks to provoke an emotional response, conflicts, animosity, or disagreements within an online social community. Trolling is a form of cyberbullying that has harmful consequences, such as causing worry and sadness, for the mental well-being of its victims. Trolling can manifest on any online platform that provides a public space for individuals to express their thoughts and viewpoints, and it can assume various manifestations:

- *Doxing* is the act of a troll revealing, on Internet, personal information about someone, such as social security number, bank account details, home location, or phone number. This puts the victim in risk of physical harm or financial loss. Doxing, although often highlighted in news reports involving politicians and celebrities, is a threat that can target individuals from all social categories.
- *Swatting* occurs when an individual engages in trolling behavior by placing a fraudulent call to emergency services, resulting in the victim being suddenly confronted by a SWAT team at their residence, leading to their arrest or causing them to experience fear.
- *RIP trolling* refers to the act of trolls deliberately posting offensive and malicious words or photographs on Facebook memorial sites dedicated to individuals who have recently passed away.

Trolls are present on social media platforms such as TikTok, Facebook, Instagram, Snapchat, Twitter, YouTube, as well as on online forums and chat rooms.

**TYPOSQUATTING.** The defining characteristic of these attacks is the utilization of email addresses or URLs that resemble an organization's domain by changing something in the name component, either replacing a letter, a number or reversing some characters, as in the examples shown below (Figure 1).

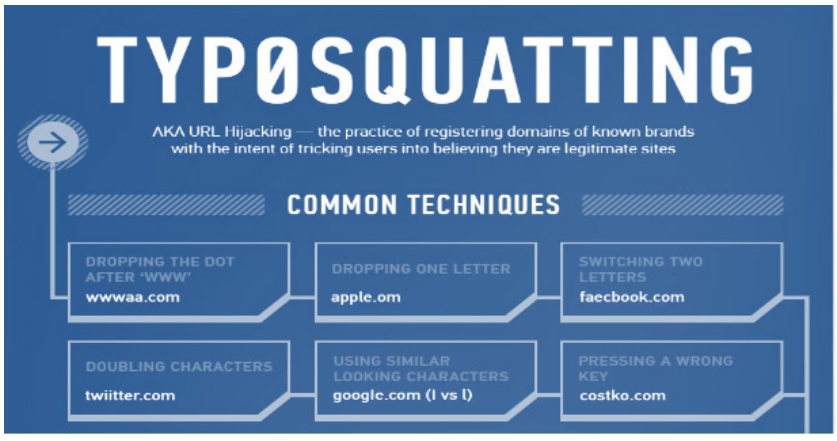


***Figure 1.*** *https://digigrow.co/2021/10/identify-and-remove-spam-from-website/*

Typosquatting attack capitalizes on the lack of attention of users who commit errors while inputting the URL of a desired website (for example misspelling Amazon as Amazun, YouTube as YouZube) and failing to notice the successive URL changes on their address bar, indicating the redirection to an unexpected website created by a cybercriminal just for such a situation (Petcu et al., 2021). One example is getting potentially unwanted applications, by inoculating the fear of a fake cyber-attack by means of a deceptive appearing notification pop up such as AMCleaner, which indicates fake system problems or failures in order to mislead Internet users into purchasing the whole edition of the application from their website (Figure 2). The application is typically obtained via Internet as a .pkg file which presents an installation graphical user interface (GUI) upon being accessed.

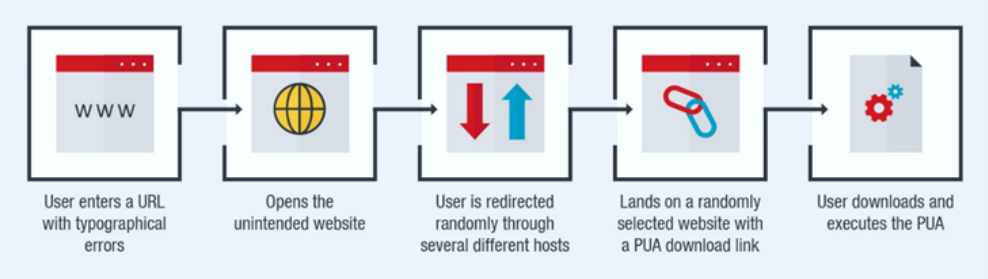


***Figure 2.*** *(Du & Magisa, 2018)*

There are various types of typosquatting, the most common being:

(1) *Bait and Switch* lures consumers into interacting with the counterfeit website which shows misleading content, such as counterfeit products, deceptive deals, or apparently harmless software downloads;

(2) *Affiliate* links imply visitors' redirection to a hyperlink that leads them to the authentic website, thus the owner of the phony site earns a commission for each redirected visitor;

(3) *Imitation Sites*: replica websites that closely resemble the original ones in appearance and design;

(4) *Malware Dissemination*: through accessing an unwittingly malevolent website;

(5) *Deceptive Surveys and Giveaways*: the phrase "Click here to give your feedback" may indicate a typosquatting attack. The fake website offers chances to "win a certain amount of money" as a means to illicitly obtain user's private information;

(6) *Website Monetization* is a counterfeit website established with the exclusive intention of making income by displaying adverts, such as pop-up ads, to the individuals that visit the site;

(7) *Joke Sites*, which are specifically designed to mock the content and features of the original website. This behavior is frequently driven by a desire for retaliation;

(8) *Phishing attacks* involve the creation of imitation websites that closely resemble real websites. The intention behind these attacks is to deceive users and trick them into revealing their private data, such as emails and account credentials.

**CLICKBAIT** is another form of attracting attention of Internet users in order to determine them to access a certain link (being either text, video or audio files) which, as a rule, presents false, sensationalized, or misleading content. This is usually done by creating teasers which raise users' curiosity, by offering a small amount of information that further on piques their interest.

From a historical standpoint, clickbait methods can be traced back to yellow journalism, which focused primarily on eye-

catching headlines with exaggerated news stories, scandal-mongering, or sensationalism, rather than offering much or any credible, well-researched news.

Clickbait headlines frequently incorporate deceit by employing convincing elements that do not adequately represent the content being provided.

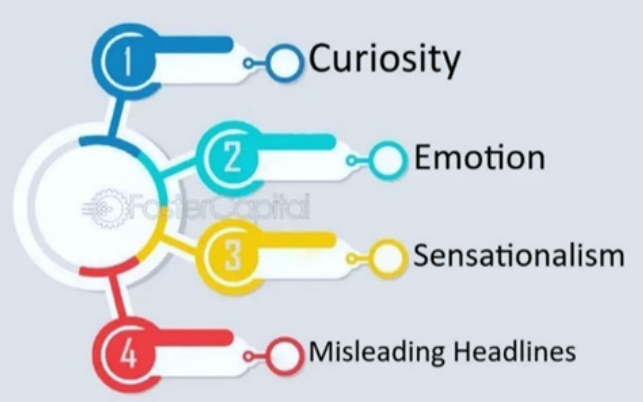


***Figure 3.*** *The psychology behind clickbait (Faster Capital, 2024)*

Most frequently, clickbait headlines include: vague language to trigger curiosity (i.e. "You'll Never Believe This"), informal language (i.e. "Salt makes you thirsty") or numerical values (i.e. "How to achieve results using 5 weird tricks" or "8 things you need to do before proposing"). The most recent clickbait headline, first circulating on TikTok, is "the POV" (point-of-view). Usually, these headlines are employed in films (such as TikToks, Instagram Reels, YouTube Shorts) where the camera tracks the main topic during travels, daily activities, facing difficult, etc. Occasionally, it is employed in a comical manner, in which the true point of view of the individual is not shown, but rather a carefully selected encounter that they wish for you to accept or follow.

## Tools designed and developed by the AI programmer:

**BOT** is an automated account, which is powered by artificial intelligence software that simulates how people might behave online. It is quite hard to remove a bot once it has entered your system. But identifying the bot itself is more of a hassle. The bots have the capacity to conceal themselves within the dark corners of the compromised system.

Bots (Computer Bots and Internet Bots) can be used digitally for **good purposes**, when gathering information for certain legitimate businesses/organizations in the benefit of the final consumer of goods and services, developing futuristic dynamic websites or making possible automatic interaction with instant messaging as well as **malicious purposes** such as gathering passwords, recording keystrokes, acquiring financial information, transmitting spam, capturing and analyzing packets, initiating DoS attacks, creating access points on the compromised machine, and exploiting access points created by worms and viruses (Figure 4).

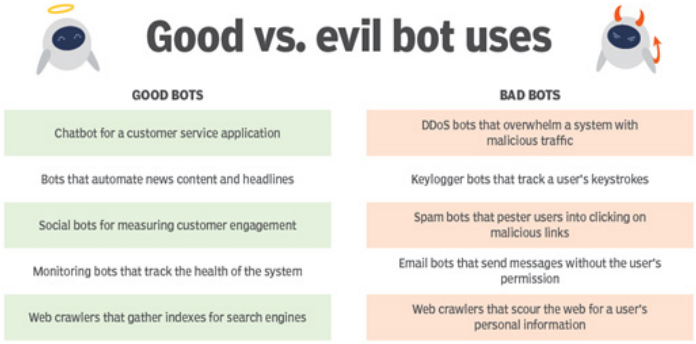


***Figure 4*** *(Lutkevich & Gillis, 2023)*

**DEEPFAKE**, a highly significant contemporary menace, involves the modification and/or fabrication of audio or video material using artificial intelligence.

Since their emergence, deep fake technologies have rapidly advanced, while production costs of high-quality deep fake material have decreased. Moreover, according to predictive scenarios, based on current trends, by 2030, the generative AI market is projected to reach a valuation of over $100 billion, with an average annual growth rate exceeding 35 %. These extreme exponential advancements in generative AI make the task of spotting deep fakes even more difficult.

To give a clearer picture of the present reality, it is worth mentioning the statistical data presented by Sensity.ai (a visual threat intelligence company from Netherlands) which claims that, since 2018, the quantity of faked videos on Internet has dramatically increased, nearly tripling every six months. This represents a very serious threat also since many AI researchers state that, with the help of AI, deep fakes will become undetectable, with its dangerous subversion of anything that is factual. With extensive training on large amounts of visual data, modern AI models can effortlessly generate new visuals that are nearly indistinguishable from actual ones. This is because pornography serves as the primary attraction for deep fake creators. The term "deep fake" came in late 2017 from a person who released pornographic movies on Reddit. These videos featured the faces of Hollywood superstars including Gal Gadot, Scarlett Johansson, and Taylor Swift, which were digitally transposed onto existing explicit clips.

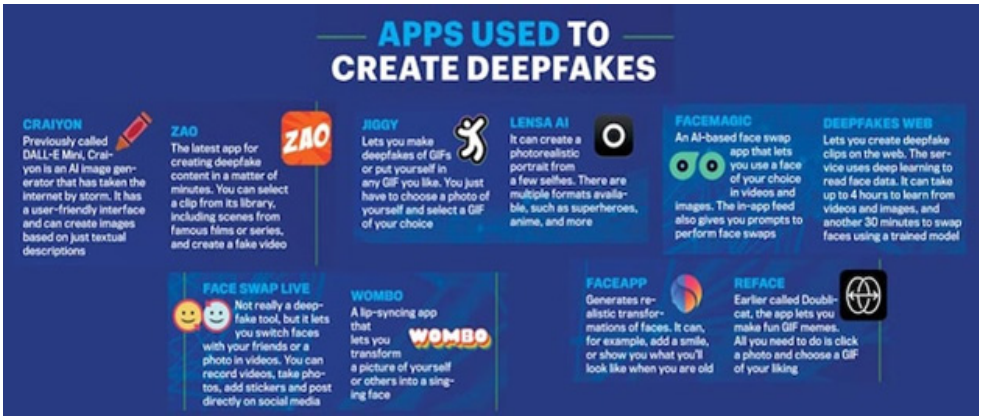


***Figure 5*** *(Sukumaran, 2023)*

So, since the challenging age of deep fakes has already begun, attention MUST BE DRAWN to the legislators in terms of protecting the rights of people whose likeness and voice are being copied. Moving forward, people must realize that, in the wrong hands, deep fakes pose a threat to national security itself, because it can also influence the results of elections, stoke social unrest, create internal or even diplomatic conflicts, disrupt public order, etc.

## THE MOST EFFECTIVE DIGITAL TOOLS TO FIGHT ONLINE DISINFORMATION AND FAKE NEWS

Reuters Institute for the Study of Journalism conducted Digital News Report (2023) revealing that Romanian news users express apprehension regarding their capacity to differentiate between fake news and authentic news on the Internet. Furthermore, of all respondents, just 48% were aware of what fake news was, and 22% declared that fake news cannot be shown on television or in news outlets.

In this framework, it is easy to understand for any novice in the field that awareness of the dangers and impact of fake news and disinformation among Internet users, of all ages, must be taken into account at the level of public policy.

That is why the responsible authorities in the field (the responsibility being shared in this case between several institutions and public authorities) must show evidence of a more responsible and results-oriented involvement than they have had until now. More concretely, the measures must be adapted to the present time and to the scale with which this phenomenon is spreading on a global level. A paradigm shift is needed, starting from the launch of aggressive awareness campaigns targeted towards the population of all ages, distributed through all mass communication media channels, especially in the context of the creation of the citizen's digital identity (eID) (Radu & Petcu, 2021), in parallel with the introduction of a separate discipline or the inclusion of this topic as a branch of a discipline such as ITC in the education system, because education, in terms of identifying and combating fake news, must start at an early age, in an increasingly digitized world.

This should happen especially in the context of some research showing that elderly adults typically exhibit greater levels of fake news awareness compared to the younger generations. Adolescents and young adults, who have a higher level of engagement with digital media, may encounter a larger quantity of false information and deceptive news articles (Pennycook & Rand, 2019).

The measures taken into account must also target the training of public administration personnel in terms of identifying fake news in the context of increasingly rapid developments regarding the digitalization of public administration. It is necessary to co-opt the private environment for the contribution of know-how in the field of reference, collaboration with specialized personnel such as experts in data mining and analysis to process relevant data, investing in advanced digital technologies and protocols for assessing content, such as specialized software to mine, organize and aggregate vast amounts of digital data.

One's high level of fake news awareness, meaning possessing knowledge on the recognition of the existence and prevalence of false, misleading content in the media landscape and the skill to distinguish between genuine and false news, leads to mitigating the impact of false news on society.

### Basic evaluation techniques

According to one of the most well-known rules of professional journalism, not every information can be taken as such and turned into news. To be a documented, complete news, it must answer six essential questions (who?; what?; when?; where?; why?; how?).

Among the basic requirements that must be taken into account for information to become news are accuracy, coherence, brevity, clarity. It

is important to consider that an article or news report should accurately convey information, include references to sources, mention all relevant parties, avoid biased perspectives, and use neutral language when quoting sources in accusing statements. Of course, there are also additional clues in evaluating the credibility of a news item in the online environment. First the title of the respective article/news/post can be taken into consideration and, if it promises the sensational, then it is a first red flag. Then, the attention should be turned to the author's writing style: if he uses evasive wording like "according to a study", "it was stated that", "in a recent press conference", "a famous analyst drew attention on the danger", most of the time, the author aims to lead to the propagation of rumors, false information, the induction of erroneous assumptions, the imprinting of wrong beliefs, in some situations even indoctrination, but also to induce a variety of feelings such as hatred, terror, drama, empathy, etc., depending on a given agenda.

At the same time, the issue of web design of the respective address should not be excluded either, which can also provide clues; more concretely, a professional design, in which time and money have been invested, shows greater credibility.

By utilizing the provided criteria and requirements while maintaining a skeptical approach towards sensational or alarmist news, and by employing a critical and investigative mindset when evaluating articles or news, individuals can establish a strong foundation for identifying online fake content. This approach also offers an advantage in the ongoing efforts to familiarize with the techniques for identifying and combating fake news and disinformation.




***Figure 6.*** *Main steps on how to spot fake news (Mind Tools, n. d.)*

## Advanced evaluation methods

Identifying false information requires examining different forms of data, including written or visual content, social context, and network structure.

In the following, by way of recommendation, but still not exhaustively, the present work presents some technical approaches and solutions that can be utilized by Internet users to detect and filter out sites/online sources containing false and misleading information.

By far, the most effective way to identify fake news and disinformation is through the use of advanced *Machine Learning Techniques (MLT)* which are crucial in handling the immense amount of information shared on social media platforms. MLT is a subfield of AI that specializes in creating algorithms and models that enable computer systems to learn and enhance their performance through experience, without the need for explicit programming. Machine learning techniques are specifically developed to analyze and extract patterns from data, enabling the system to identify and comprehend intricate relationships and make informed predictions or decisions. Various machine learning algorithms are utilized based on the type of data.

Mainly, MLT encompass the following:

1) *Deep Learning Techniques* – are cutting-edge techniques that utilize artificial neural networks with multiple layers to gain insights and recognize patterns in data, identifying and flagging fake news through the analysis of linguistic features present in news articles;

2) *Natural Language Processing Technique (NLP)* – is based on an algorithm that allows computers to efficiently process, analyze, and comprehend human language. These techniques cover a broad spectrum of tasks, including text classification, information extraction, sentiment analysis, machine translation, and question answering. Working with NLP techniques requires the utilization of statistical models, machine learning algorithms, and linguistic resources to address the complexities and ambiguities present in natural language;

3) *Ensemble Learning* – through the combination of various models, the overall performance can be enhanced, often exceeding the capabilities of any single model. Ensemble Learning has proven its efficiency in detecting fake news, by combining the outputs of various algorithms, resulting in an enhanced detection accuracy. It has the expertise to address vulnerabilities in specific models and minimize any potential biases;

4) *Transfer Learning* – enables the utilization of pre-trained models for novel tasks. It has been utilized for identifying fake news through the application of pre-trained models for natural language processing tasks and refining them specifically for fake news detection. Transfer Learning is a valuable technique that enhances a learner's performance by leveraging knowledge from a related domain. By examining real-life situations outside of the technical realm, insights into the feasibility of Transfer Learning can be gained. Acquiring training data that accurately reflects the feature space and projected data distribution of the test data can be challenging and costly in specific situations;

5) *Graph-based Techniques* – allow for the analysis of connectivity patterns, community structures, and network properties. Graph-based Algorithms are a popular technique for representing news articles as graphs and analyzing the relationships between nodes/ vertices connected by edges in order to identify fake news. These algorithms have demonstrated encouraging outcomes in detecting fake news.

In general, machine learning is a valuable tool for identifying and stopping the dissemination of false information on social media.

For instance, reliable news sources use the same domain name as the source; shady ones use typosquatting to hide this. In this case, it is rather simple, as even the simplest precaution is crucial in safeguarding against typosquatting threats. To avoid landing on these rogue pages, users must pay attention when typing the URL. For example, the country of origin (.ro, .de, .at, .uk) or the purpose (.eu, com,. edu) are indicated by the URL's ending and the links within news items. One can also verify who owns that domain by accessing https://www.whois.net. Furthermore, since fabricated images frequently bolster false information, two alternatives for discerning comparable or visually identical web images can be recommended: CLOUD Vision API, namely the Google Images algorithm, and Tin Eye Reverse Image Search, a service that provides image search solutions to a broad spectrum of users.

Besides, a browser extension known as *Trusted Times* categorizes unreliable and false news. It provides additional information about articles, including bias and details about who is positively and negatively covered, through the use of machine learning. It is capable of categorizing articles as false, unreliable, verified, or mainstream media, summarizing verified content, identifying significant topics, detecting potential bias, determining the historical bias of reporters, and determining the historical bias of a particular website. It applies icons to demonstrate the trustworthiness of websites.

For example, Clickbait Headlines detection can be successful by making a clickbait classifier in Python, using an application of the *OSEMN* data science processes along with *NLP* and *Machine Learning* algorithms.

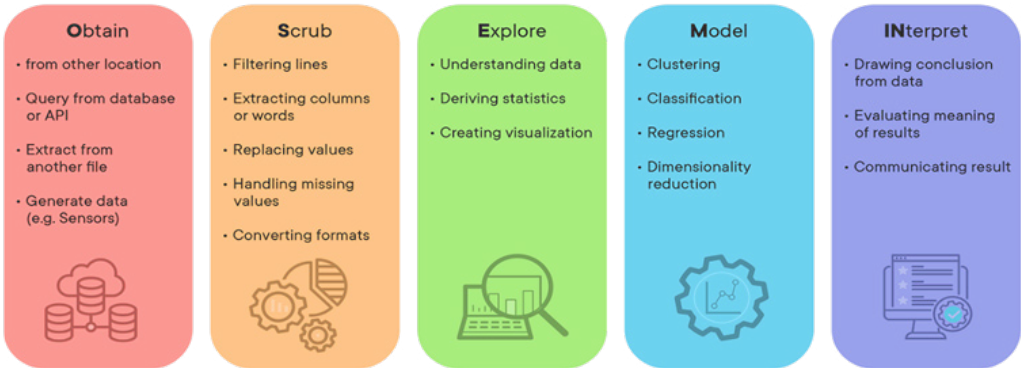


***Figure 7.*** *Data Science ASEMN Model (Kim, 2021)*

Additionally, users should download files or software only from official websites or trusted sources, in order to prevent encountering such dangers.

An effective method, albeit costly, to safeguard against typosquatting of owners of websites is to secure trademark for their domains and acquire all associated domain names that are prone to misspelling. If a domain has been trademarked, it is indeed possible to file complaints against individuals who engage in typosquatting.

## How can a user know if he is being trolled?

There are several distinct indicators that suggest someone may be encountering a troll. Trolls often try to target something for which the user has openly shown affinity or support. If someone is deliberately attempting to denigrate the user's preferences or engaging in a general derogatory discourse about a forum or post, it is quite likely for that person to be dealing with a troll. Trolls are notorious for making extravagant and audacious assertions. They frequently make absurd assertions regarding the topic being discussed, with the intention of provoking an emotional reaction from others.

Trolls may not genuinely believe the outlandish assertions they make, but their goal is to convince others. One may identify a troll by their tenacity. Trolls persist in their actions until they have successfully provoked a strong emotional reaction from their target.

## How can malicious bot activity be prevented?

It is well-known that preventing malicious bots is crucial in maintaining strong security measures. Implementing strong cybersecurity measures can effectively prevent bot infections. Here are some methods to prevent bots:

1. Make sure to install antimalware software. Malicious bots are classified as malware. Utilizing antimalware software can automate the process of safeguarding against this particular type of threat;
2. Set up a bot manager. A bot manager is commonly found within a web app security platform. These categories web requests and enable the utilization of certain bots, while preventing others from accessing. There are two strategies for managing bots:
   - Rate limiting restricts the number of requests from the same address;
   - CAPTCHAs employ puzzles to confirm the authenticity of the user making the request, distinguishing humans from bots.
3. Make sure to utilize a firewall. Firewalls have the capability to block bots and restrict specific traffic by analyzing IP addresses or behavior;
4. Keep the software up to date. Software updates include important security updates that can help protect against malicious bots;

5. Remember to practice good *password hygiene*. Software engineers are well aware of the potential risks associated with bots that can be used to forcefully crack weaker passwords and gain unauthorized access to user accounts. Ensuring a robust password and regularly updating it can be effective in mitigating this issue;
6. Only click on *links that* you trust. Emails can be used by bots to send spam or malicious links. Make sure to only click on links from a reliable source to steer clear of any potentially harmful links from automated **programmers**.

Since well-crafted spear phishing attacks are now an everyday occurrence, they can be combated by using Threat Intelligence tools. Antimalware software agents must also be installed on computer systems. Another preventive measure is to ensure IDS/IPS rules are set for associated malware.

## How to detect deepfake?

As it has been already clarified in the previous dedicated subsection, detecting deep fakes can be challenging as technology advances. According to a report from KPMG, there are over 100,000 AI-based models that can be used to generate or identify deepfakes. Out of these, only around 3,000 have the ability to accurately detect deep fakes, with most of them only able to identify the less sophisticated ones.

For example, there are a few available ways to detect deep fake videos, such as:

- *Visual inspection:* Occasionally, deep fakes may exhibit artefacts or anomalies that are absent in genuine videos. These artefacts can manifest as unusual flickering, distorted visuals, or out-of-sync lip movements;
- *Machine learning:* Another approach involves utilizing machine-learning algorithms that can be trained on a vast dataset of authentic and fabricated videos. These algorithms can then be employed to categorize new videos as either genuine or counterfeit;
- *Metadata analysis:* It's fascinating how digital files can hold valuable metadata that allows for source tracking and authenticity verification. Through careful analysis of a video's metadata, it is possible to ascertain if any editing or manipulation has taken place.
- *Forensic techniques:* When it comes to analyzing videos and detecting deep fakes, there are a variety of forensic techniques available. These techniques could involve analyzing patterns in the video, examining the audio track, or comparing the video to other sources to identify any discrepancies;
- *Employing various techniques:* Using a combination of methods is the most effective approach to detect deep fakes. Through the integration of visual inspection, metadata analysis, forensic analysis, and machine learning, a more precise evaluation of a video's authenticity can be achieved.

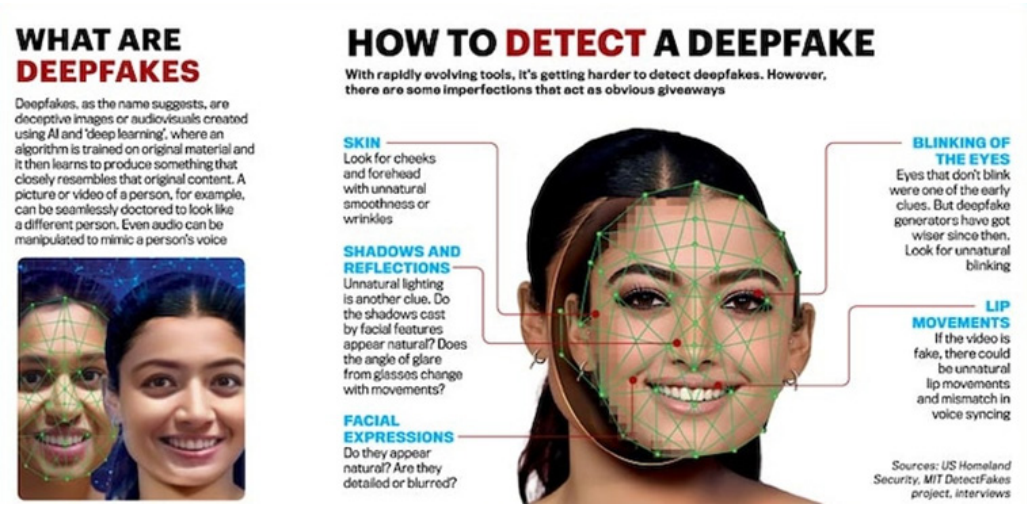


***Figure 8.*** *(Sukumaran, 2023)*

## CONCLUSION

The transnational spread of both disinformation and misinformation can lead to a variety of detrimental outcomes, including endangering democratic systems, intensifying divisive discussions, and jeopardizing the well-being, safety, and environment of citizens, Internet users. While propaganda has been existing for a long time, the rise of digital "fake news" poses a significant danger, mainly because of how easily content can be created, spread, and consumed. Social networking platforms, Internet users, websites, and blogs all contribute to the problem to varying degrees. In the present digital landscape, various individuals and organizations, including political factions and extremists, have the ability to create successive waves of fabricated information, commonly known as fake news. These falsehoods have the potential to have significant influence on public sentiment, ultimately shaping both global and local agendas. By employing predictive algorithms, these perspectives can be disseminated online, causing the content to become viral specifically to persons who are most inclined to propagate and reinforce such themes. Hence, it is crucial for proficient governmental bodies to comprehend and effectively tackle the perils of false information, misinformation, and disinformation. They should enhance their response by fostering public awareness, collaborating with specialized personnel, such as data mining and analysis experts, to handle pertinent data, and investing in cutting-edge digital technologies and protocols for evaluating content, such as dedicated software for extracting, organizing, and consolidating extensive amounts of digital data.

So, given the extent of the danger to which users of the online space are exposed in terms of fake news, disinformation and the digital means by which this information is disseminated, as it has been mentioned above some solutions meant to combat this danger were outlined, namely: on one hand, raising public awareness on the implications of fake news and disinformation, informing online users about the existing tools for launching and distributing fake news, disinformation and misinformation and the easiest and most efficient methods to identify and combat them, and, on the other hand, working with specialized personnel such as experts in data mining and analysis to process relevant data, investing in advanced digital technologies and protocols for assessing content, such as specialized software to mine, organize and aggregate vast amounts of digital data.

Undoubtedly, through extensive exploration of machine learning techniques such as data mining, deep learning, natural language processing, ensemble learning, transfer

learning, and graph-based methods, researchers have made remarkable strides in detecting and addressing the issue of fake news on social media platforms.

Even if valuable progress is undeniable and ongoing, there are still limitations at this point, such as:

- *Social media data* mining provides valuable insights into user behavior and network dynamics, but encounters difficulties related to data quality, data availability, and computational demands;
- *Deep learning* can learn complicated patterns and semantic knowledge, although its effectiveness is constrained by the scarcity of annotated data, high processing costs, and the challenge of interpretability;
- *NLP* enables the discovery of linguistic patterns and proved a profound impact on detecting fake news, but it lacks contextual information and faces challenges in dealing with textual ambiguity. Though, it is noteworthy that a new trend has emerged: leveraging quantum computing to enhance the performance of NLP models;
- *Ensemble learning* addresses the limitations of individual models, but it incurs significant computing expenses and system intricacy;
- *Transfer learning* streamlines the training process, but depends on the availability of appropriate pre-trained models and the resemblance of data;
- *Graph-based techniques* offer valuable insights into the dissemination and impact of false information, but they require the use of high-caliber data and are deficient in terms of textual context.

As the realm of fake news detection is constantly changing, researchers keep exploring various methods to enhance performance, efficiency, and interpretability. By utilizing various methods and refining strategies, including combining techniques and optimizing models, further improvements and easily expandable solutions are expected to be seen in the future. Nevertheless, various machine learning methodologies are better suited for certain types of data.

## REFERENCE LIST


Adams, Z., Osman, M., Bechlivanidis, C. & Meder, B. (November 2023) (Why) is Misinformation a Problem?. *Perspectives on Psychological Science*. 18(6), 1436-1463. doi: 10.1177/17456916221141344.

Aldwairi, M. & Alwahedi, A. (2018) Detecting fake news in social media networks. *Procedia Computer Science*. 141, 215–222. doi: 10.1016/j.procs.2018.10.171.

Berrondo-Otermin, M. & Sarasa-Cabezuelo, A. (2023) Application of Artificial Intelligence Techniques to Detect Fake News: A Review. *Electronics*. 12(24), 5041. doi: 10.3390/electronics12245041.

Bloch, M. (1921) Réflections d'un historien sur les fausses nouvelles de la guerre [Reflections of a Historian on the False News of the War]. *Revue de Synthèse Historique*. 7, 13–35.

Du, S. & Magisa, L. (25 July 2018) Typosquatting Tactic Tricks Mac Users into Downloading Potentially Unwanted Application. *Trend Micro*. https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/typosquatting-tactic-tricks-mac-users-into-downloading-potentially-unwanted-application

Faster Capital (4 April 2024) *Clickbait: Don't Take the Bait: Phishing and Clickbait Traps.* https://fastercapital.com/content/Clickbait--Don-t-Take-the-Bait--Phishing-and-Clickbait-Traps.html

Ireton, C. & Posetti, J. (eds.) (2018) *Journalism, "Fake News" and Disinformation: A Handbook for Journalism Education and Training*. Paris, France, UNESCO.

Kim, T. J. (2021) Clickbait Headline Detector using NLP and ML algorithms. *Medium*. https://taejoon-kim.medium.com/clickbait-headline-detector-using-nlp-and-ml-algorithms-23c248c945ff

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook,

G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J. & Zittrain, J. L. (2018) The science of fake news. *Science*. 359(6380), 1094–1096. doi: 10.1126/science.aao2998.

Lutkevich, B. & Gillis, A. S. (March 2023) What is a bot?. *TechTarget.* https://www.techtarget.com/whatis/definition/bot-robot

Mind Tools (n. d.) *How to Spot Real and Fake News.* https://www.mindtools.com/a0g6bjj/how-to-spot-real-and-fake-news

Petcu, I., Candet, I. B., Ştefănescu, C., Gruia, C. I. & Craioveanu, V. (2021) Security Risks of Cloud Computing Services from the New Cybernetics' Threats Perspective. *Romanian Cyber Security Journal*. 3(1), 89-97.

Pennycook, G. & Rand, D. G. (2019) Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences of the United States of America*. 116(7), 2521-2526. doi: 10.1073/pnas.1806781116.

Radu, A. F. & Petcu, I. (2021) Intrinsic aspects of e-Government consolidation across the European Union. Case study: Romania. *Romanian Journal of Information Technology and Automatic Control.* 31(4), 83-96. doi: 10.33436/v31i4y202107.

The Reuters Institute for the Study of Journalism conducted Digital News Report (2023), https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf

Social Media Victims Law Center. (26 February 2024) *What is trolling on social media?.* https://socialmediavictims.org/cyberbullying/types/trolling/

Staats, B. (11 February 2021) Misinformation, disinformation, malinformation: What's the difference?. *Minitex*. https://minitex.umn.edu/news/elibrary-minnesota/2022-05/misinformation-disinformation-malinformation-whats-difference

Sukumaran, A. (7 December 2023) Deepfakes: Clear and present danger. *India Today*. https://www.indiatoday.in/india-today-insight/story/deepfakes-clear-and-present-danger-2473187-2023-12-07



This is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License.