



Information and its Reliability in the Digital Reality

Laman Surkhay Suleyman-zada
Baku State University, Azerbaijan

Abstract

The latest scenarios involving multi-modal messaging, and the multi-faceted and dynamic nature of online communication, act as barriers to finding reliable information. Digital technologies and artificial intelligence, given the large-scale and uncontrolled dissemination of information in real time, create fertile ground for disinformation, fake news, false information, various kinds of hoaxes and manipulation used for political and commercial propaganda. Disinformation, which employs complex manipulative narratives and social engineering techniques, poses a threat to society as a whole, given its political, social and psychological dimensions. By polarising society, it can give rise to anti-democratic and populist movements and cause serious damage to democracy. The aim of this study is to examine information and disinformation as tools of propaganda, a form of information warfare and a means of ideological struggle. The specific nature of the subject matter necessitated the use of an interdisciplinary methodology, incorporating approaches from political science, sociology and mass communication theory. The methods employed in this study include observation, document analysis, comparison, analysis and synthesis. The study covers the period from the COVID-19 pandemic, Russia's war against Ukraine and the war between Israel and Hamas, up to the present day – the war between the US and Israel against Iran. The theoretical significance of the work lies in the systematisation and synthesis of expert assessments and academic articles on the subject of false information, fake news and disinformation. The current priority is to devise effective measures to combat the spread of disinformation online, develop a fact-checking system, and establish the credibility of information sources. In conclusion, the paper argues for the need to establish independent international media outlets to disseminate unbiased information on digital platforms.

Keywords: Disinformation, Fake News, Misinformation, Propaganda, Information Warfare

1. Introduction

The imbalance in today's digital information landscape means that the average user is less able to critically filter information amidst the flood of disinformation, false and malicious content that dominates news feeds and official websites. Communication platforms have become bridges between media corporations – as providers of information – and users, driving large-scale user interaction and commodifying human attention. The economic model

of ‘clickbait media’, in which traditional news organisations rely on monetising click-throughs, carries no obligation regarding the accuracy of the information published. (Qorib et al., 2024). The resulting media landscape fosters extreme content, which leads to a toxic culture of interaction across all social media platforms, as it is precisely this content that attracts vast amounts of capital to the internet (Сулейман-заде, 2024). According to the results of the Reuters Institute’s 2024 study, content with emotionally charged headlines (‘threat’, ‘scandal’) spreads 2.3 times faster than neutral content. (Digital News Report, 2024)

The accuracy and objectivity of information are always important, but take on particular significance during wars, armed conflicts and crises, when the public uses the online space to collectively make sense of the situation. With every global event, there is growing concern about the fragility of truth in public discourse and how a culture of dishonesty undermines the social relationships that ensure the stable functioning of democracies. This was the case during the COVID-19 pandemic, the war between Israel and Hamas, and the current conflicts involving Russia against Ukraine, the US, and Israel against Iran, when manipulation and disinformation in the information sphere fuel socio-political tensions, provoke protest sentiments, and pose significant threats to national security.

In 1992, in an essay ‘The Government of Lies’, devoted to events in the Persian Gulf, the Serbian-American playwright Steve Tesich, expressing scepticism about how the government of George H. W. Bush had presented the reasons for the conflict to the public and, crucially, how they had been perceived by the public, remarked that the era of ‘post-truth’ had dawned (Tesich, 1992). Almost two decades later, in November 2016, the term ‘post-truth’, according to the editors of the Oxford University Dictionary, was named ‘Word of the Year’ due to two political campaigns – the United Kingdom’s exit from the European Union and the US presidential election. The accompanying definition was as follows: ‘describing circumstances in which objective facts are less important in shaping public opinion than appeals to emotion and personal belief’. Since then, the phenomenon has become firmly entrenched in politics and the mass media, giving rise to ‘post-truth politics’, which has sparked opportunistic, speculative narratives and emotional, unfounded appeals to public opinion. Drawing on Nietzsche’s thesis that there are no facts, only their interpretation, experts describe the phenomenon as a crisis of epistemic authority, in which facts lose their significance not only due to manipulation, but also due to indifference towards generally accepted standards of justification.

Technological innovations—ranging from the printing press, radio and television to the internet—have influenced the scale and speed at which information is disseminated, with much of it now transmitted in real time. Digital technologies and artificial intelligence play a major role in shaping the modern media landscape, as well as in the devaluation of accurate and verified facts. In his article ‘Freedom of Speech in the Algorithmic Society...’ (Balkin, 2017), Jack M. Balkin argues that in the era of the algorithmic society, ‘artificial intelligence agents’ and ‘Big Data’ are capable of controlling the population through digital infrastructure, large multinational social media platforms and search engines, which stand between traditional nation states and ordinary people and act as special regulators of public opinion. According to the author, the algorithmic society gives rise to two main problems concerning freedom of expression: firstly, big data enables new forms of manipulation and control, which private companies will attempt to legitimise and shield from regulation by invoking the principles of freedom of speech; secondly, social media and online platforms currently shape people’s consciousness in much the same way as nation-states once did.

Social media ecosystems and artificial intelligence have made it easier to create and widely disseminate fabricated information, fake news, blatant propaganda, manipulative content and

disinformation to a broad global audience, and are in fact ‘used as a testing ground for new methods of informational and psychological influence on large population groups and the construction of a desired reality. The means and forms of transmitting specific information constitute ‘ensembles’ of communicative practices and produce new meanings, whereby we are dealing with a multiplicity not only of truths, but also of subjects claiming the right to proclaim these truths.

The manipulation of information and its impact on the formation of public opinion constitute a little-studied area—a phenomenon requiring in-depth and fundamental political and philosophical analysis. Experts identify three levels of manipulation of public opinion (Сулейман-заде, 2024):

- A change in the values, ideas or attitudes already present in the public consciousness;
- A change in attitudes towards a significant event in society;
- A complete shift in society’s attitudes towards life.

As far back as 1945, Edward Bernays, in his book «Public Relations», referred to the emerging science of shaping public opinion as the ‘engineering of consensus’(Bernays, 2013). For Bernays, shaping public perception was the foundation of the democratic process of governance: ‘Those who manipulate this invisible mechanism of society constitute an invisible government, which is the true ruling power of our country... The conscious and intelligent manipulation of the habits and opinions of the masses is an essential element of a democratic society.’

Today, TikTok, X and Meta have the capacity to disseminate information—whether true or false—instantly, thanks to algorithmically curated feeds, high speeds and a sense of anonymity (Steinebach, 2020). Even after being debunked, fake posts still have the desired effect, becoming more popular as they are widely discussed. Neural networks, now accessible to the average user, can generate plausible text that is sometimes indistinguishable from that written by a human. Deepfake technology has made it possible to create fake news about politicians by generating images, videos and voice recordings. Disinformation has become a global problem, eroding institutional authority and undermining trust in official information. Spreading across social media, it is often amplified by influential figures, 62 % of whom, according to UNESCO’s large-scale 2024 study ‘Behind the Screens’, do not verify information before sharing it. (UNESCO, 2024)

The theoretical underpinnings of this research are based on an analysis of research articles from influential scientific and political journals, academic databases and reports from well-known companies. When examining facts from primary sources, their validity was verified in at least two authoritative academic resources.

Object of the study. The object of the study is information and its dissemination in the digital space.

Focus of the study. The focus of the study is the ecosystem of digital communication platforms as a source of information and disinformation in the contemporary media landscape.

Aim of the study. To examine information and disinformation as tools of propaganda, a form of information warfare and a means of ideological struggle.

The aim of this study is to put forward proposals for developing effective measures to verify information and prevent disinformation on the internet.

Research hypothesis. It is hypothesised that, given the political, social and psychological aspects of digital disinformation, its use as a high-tech weapon within the context of hybrid conflict enables geopolitical objectives to be achieved without the need for overt military action.

2. Methods

The specific nature of the subject under study led to the adoption of an interdisciplinary methodology, incorporating approaches from political science, sociology and mass communication theory. The research methodology is based on an understanding of an innovative communication system, with its spatial-temporal, collective-individual and subject-specific dimensions, drawing on the achievements of sociology, political science and media psychology; on an understanding of the interrelationship between the technological, sociocultural, political and media spheres, as well as the discursive construction of reality and communicative subjects, and their reflection in contemporary online practices of information exchange.

The methods used in this study were: observation, to identify manipulative techniques in media streams on YouTube, TikTok, X and Meta; analysis of documents, including scientific articles and papers, official statements by politicians, and social media posts; comparison and analysis of the spread of disinformation and manipulative information during the period covering the COVID-19 pandemic, Russia's war against Ukraine, the war between Israel and Hamas, and the war between the US, Israel and Iran; synthesis to produce a final classification of contemporary techniques and types of disinformation and manipulative information.

3. Results

At the present stage, disinformation, manipulative content and political propaganda are closely intertwined with digital media. They are driven by state and non-state, political and commercial actors, the media, citizens and groups of citizens — as well as by the use of communication infrastructures employed to produce, disseminate and amplify disinformation on a large scale. The global 'post-truth' society is characterised by a high concentration of false messages, propaganda machines, biased algorithms, malicious artificial intelligence applications, social media bots and misleading information campaigns.

Digital disinformation is a key feature of contemporary international relations; however, the manipulation of information, propaganda and disinformation are not the preserve of any single country, political regime, nor are they a typically modern phenomenon — they are classic instruments of foreign policy employed by many states, with a history stretching back almost a thousand years, to the time of the first Roman emperor, Octavian's, smear campaign against Antony, with its 'short, sharp slogans inscribed on coins in the style of archaic tweets'.

In 2018, the European Commission's Expert Group on Fake News and Online Disinformation provided the following definition: 'Disinformation encompasses all forms of false, inaccurate or misleading information that is created, presented and disseminated with the intention of deliberately causing harm to the public or generating profit.' (Freelon & Wells, 2020)

In the author's view, the classic definition of propaganda was provided by Harold Lasswell: 'the expression of opinions or actions by individuals or groups with the aim of deliberately

influencing the opinions or actions of other individuals or groups in order to achieve predetermined objectives’.

Disinformation is characterised by the falsification of data, a potential threat and a destructive intent. Propaganda and disinformation, as targeted informational influences on public consciousness for political ends, have become the basis of interstate confrontation and military information and psychological operations at both the tactical and strategic levels. Disinformation campaigns directly affect national security issues; as Francis Fukuyama asserts, the use of ‘bad information’ is a weapon for subduing the enemy in information warfare campaigns. (Fukuyama, 2021). Thomas Ron, an adviser to the US Department of Defence, in 1976, was the first to define as ‘strategic, operational and tactical competition at all levels (from peacetime to open conflict), in the course of which adversaries use information assets to achieve their political, military and economic objectives’ (Kuehl, 2002).

Disinformation, as a foreign policy tool, may form part of a much larger and more dangerous complex of international state activity in the cyber domain, including cyberattacks, hacking and other subversive activities, collectively referred to as ‘hybrid warfare’. It is capable of having a profound impact on international relations: fuelling mutual mistrust, bringing conflicts into the public sphere, creating tension in diplomatic relations and provoking armed conflict. Awareness of the risks of manipulating information for political ends has risen sharply following repeated foreign interference in the domestic political processes of sovereign states, which directly affects issues of their digital sovereignty. (Сулейман-заде, 2025)

Disinformation can be fabricated and disseminated by various actors, and can take many forms and types: a disinformation narrative, a disinformation campaign, or a disinformation operation. It is difficult to trace the sources, problematic to identify the actors involved, and even more so to determine the attribution behind the false content. The perpetrators of disinformation may include both domestic and foreign forces, various political actors, agitators, hacktivists and terrorists who use disinformation campaigns to manipulate public opinion and undermine trust in the country’s institutions. (Сулейман-заде, 2025)

To analyse the spread of disinformation and manipulative information in the digital society and their impact on it, the author examines the period encompassing the COVID-19 pandemic, Russia’s war against Ukraine, the war between Israel and Hamas, and the war waged by the US and Israel against Iran. It was precisely between 2020 and 2026 that a global transformation of the information landscape took place, laying the foundations for digital disinformation and reflecting the evolution of disinformation from spontaneous information cascades to full-scale, coordinated operations driven by artificial intelligence algorithms and neural networks.

The COVID-19 pandemic has acted as a catalyst for the spread of misinformation or information taken out of context. Typical scenarios have included rumours of planned campaigns for compulsory vaccination with a secret drug and discriminatory measures against people from China. Since the outbreak began, there has been widespread discussion online of the theory that the new virus was developed at the Wuhan Institute of Virology and entered the city as a result of an accidental leak. As COVID-19 has spread across the globe, misinformation on social media has become part of many people’s lives, leading to the emergence of a negative emotional climate, ‘waves’ of negative attitudes towards the measures taken by the authorities, as well as the proliferation of various conspiracy theories about the infection, its nature, government policy, and the strategies and tactics of public health authorities and regulatory bodies. An analysis of the phenomenon of fake news in the healthcare sector has shown that infodemic information can cause psychological distress,

panic, fear, depression and fatigue. It is noteworthy that the nature of COVID-19-related disinformation largely depended on the specific characteristics of a given country: whilst people in the West actively discussed the possibility of infection via 5G networks, in Russia the threat was perceived to lie in parcels from AliExpress. In response to the widespread dissemination of false information on social media – which the World Health Organisation termed an ‘infodemic’ – online platforms promoted content from official sources and removed advertisements for medicines and COVID-19 treatments with unproven efficacy. The COVID-19 pandemic, as the most significant challenge of 2020, spurred research into the algorithms governing the emergence and spread of fake news, the psychological aspects of manipulating public opinion (and mechanisms for countering this influence), and also raised once again the issue of fostering media literacy among the population.

At a strategic level, war is about propagating narratives about who is the victim and who is the aggressor, according to RAND experts. The conflict between Israel and Hamas, which escalated into open military confrontation in October 2023, coincided with a leap forward in the development of artificial intelligence. In the Palestinian-Israeli conflict, “where emotions run very, very high”, as Hani Farid, a professor of digital forensics at the University of California, Berkeley, noted, disinformation – including images generated by artificial intelligence – is particularly effective. The confrontation between the sides has become the ideal environment for the production and dissemination of fake content and the stoking of emotions. The realisation that AI-generated content is circulating online leads users to feel uncertain about everything they encounter.

Hani Farid explains: “The ability to manipulate photographs, audio and video material means that everything is called into question.” Sometimes, for example, genuine images are mistakenly labelled as inauthentic (Eisele & Steinwehr, 2023). At the same time, social media researchers have highlighted the large amount of disinformation that has appeared on X (formerly Twitter) against the backdrop of the war between Israel and Hamas. For instance, within a few hours of the attack on Israel, unverified photographs and videos of missile strikes, destroyed buildings and scenes of violence began to appear on X. Many of these posts used old images of armed conflicts and footage from video games, whilst some ‘right-wing experts’ even circulated viral hoaxes on the platform. Information spread online claiming that ‘the US is sending \$8 billion in military aid to Israel’, accompanied by a screenshot of a White House statement as ‘proof’. However, the screenshot turned out to be a copy of a White House statement from July regarding financial aid to Ukraine, with a falsified date. In the first three days following the attack on X (Twitter) alone, more than 50 million posts about the conflict were published, as officially recorded by X Safety.

Despite the fact that the social media platform has stated that it is removing newly registered accounts linked to Hamas and is endeavouring to prevent the spread of ‘terrorist content’, and has also removed anti-Semitic comments and hundreds of accounts attempting to manipulate public opinion (Goswami, 2023), according to Mike Rothschild, who studies the spread of viral disinformation on social media, news of the attack on Israel became “the first real test for Elon Musk’s version of Twitter, and it failed spectacularly”. “Musk’s changes didn’t just render X useless during a crisis. They actively made the situation even worse,” Bloomberg quoted him as saying.

Researchers from the independent international team Bellingcat have identified several key techniques used to spread fake images and videos:

- The use of old photographs of shelling — both in Palestine, including the Gaza Strip, and in Syria, Ukraine and elsewhere.;
- The use of old footage that does not depict any conflicts — for example, one of the videos, purportedly showing Hamas militants landing at the venue of a music festival, was in fact filmed in Egypt and posted on TikTok in September 2023; and the video of a girl allegedly being burnt alive at the festival was filmed in Guatemala in 2015;
- The dissemination of completely false reports — for example, that the Church of Saint Porphyrius in Gaza had been shelled — representatives of the church confirmed via their official account that it had not been damaged;
- The circulation of fake documents — for example, claiming that the US has approved \$8 billion in military aid to Israel.

Whilst Israel mainly waged its information war on X (Twitter), Facebook and Instagram, Hamas and Islamic Jihad used Telegram for similar purposes, as it has less stringent moderation, allowing militants to broadcast graphic videos, documenting their violent incursion into Israeli territory, the taking of hostages, the deaths of tens of thousands of innocent Palestinians in Gaza, and the destruction of the Strip's infrastructure. It is worth noting that false claims, AI-generated videos and old footage of fighting were used in the conflict by both state propaganda and individual content creators, with the aim of boosting click-through rates and monetising content.

A distinctive feature of the information dimension of the confrontation between Russia and Ukraine is that it constitutes a separate component and, in most cases, is no less important than the military conflict. In the Russian-Ukrainian 'war of words' or 'war of narratives', social media users are not merely the targets of destructive influence, but active drivers of the information war. Posts by ordinary users of digital platforms create narratives; these narratives shape political perceptions or identities, which act as driving forces behind foreign policy, undermining the power and influence of states over the course of global politics. (Сытник, 2022)

The Russian-Ukrainian conflict has become the 'first war on social media' (Suciu, 2022) and a 'testing ground' for evaluating the methods and tools of information warfare (Mölder & Sazonov, 2018). The emergence of generative language models, such as ChatGPT, has contributed to the widespread dissemination of content generated by fake accounts in the Russian-Ukrainian sphere, which is crowding out alternative narratives. Such technologies have become an effective addition to the numerous armies of bots carrying out various tasks on social media. By posing as ordinary users, they create the illusion of genuine public opinion. This high-tech weapon is used in information and psychological warfare, which have become the new reality of the modern world. Furthermore, in the Russian-Ukrainian information war, state-sponsored trolls—motivated by political, ideological and, at times, financial considerations—are deployed to shape online narratives.

The war between the US and Israel against Iran has clearly demonstrated the extent to which AI-generated videos can influence perceptions of events during periods of peak news consumption, when the countries involved in the conflict seek to shape their own narrative. The widespread affordability of AI-powered video-creation technologies since the start of the war with Iran has led to the large-scale circulation on social media of fake, AI-generated 'deepfake' videos and photographs of combat operations, and the consequences of strikes on civilian infrastructure, exacerbating disinformation and significantly reshaping public attitudes towards the ongoing armed conflict. Examples include videos purportedly showing the US aircraft carrier USS Abraham Lincoln ablaze at sea. The potential exploitation of

synthetic content through the presentation of AI-generated audiovisual material — audio, images, text, video or combinations thereof — as documentary evidence of events or statements by destabilising actors for political or economic purposes is recognised as a security threat (ENISA). Alongside AI-generated footage of combat, reports of the deaths of Israeli Prime Minister Benjamin Netanyahu and Iran’s Supreme Leader Mojtaba Khamenei circulated widely. Unverified content reaches millions of users in a matter of minutes, and the public is left to fact-check the material themselves – material that often looks as realistic as possible and is shared across several platforms simultaneously.

Disinformation campaigns are also being actively waged against Azerbaijan, both during the 44-day war with Armenia in 2020 and in the current conflict involving the US, Israel and Iran. In particular, an operation aimed at politically destabilising our country, with the aim of advancing Iran’s regional interests and undermining Western influence, was highlighted in the ‘Adversarial Threat Report, First Half 2026’, which noted that the American technology corporation Meta had, in the first quarter of 2026, removed 367 Facebook accounts, 3 pages and 360 Instagram accounts linked to Iran and targeting Azerbaijan for breaching the policy against Coordinated Inauthentic Behaviour.(Meta,2026) As stated in the report, the aim of the campaign was to introduce narratives into the comments sections of popular news sites, such as the BBC and Radio Liberty.

Table 1. Narratives into the comments sections of popular news sites

Conflict/Event	Main tool	Mechanism of influence	Main effect
COVID-19	Fake news in the healthcare sector	Fomenting uncertainty and fear	A study of algorithms for the psychological manipulation of mass consciousness
Russia–Ukraine	The widespread dissemination of content generated by fake accounts	The First Social Media War	Creating the illusion of genuine public opinion
Israel–Hamas	Manipulative content created using AI	Information chaos	The spread of ‘terrorist content’ and anti-Semitic sentiment
US–Israel against Iran	The instrumentalisation of synthetic content through the presentation of AI-generated audiovisual objects	The impact on the perception of events during periods of peak news consumption	A shift in public attitudes towards the current armed conflict
Azerbaijan	The incorporation of narratives into the comments sections of mainstream news websites	The legitimisation of political decisions under the guise of a ‘grassroots’ initiative by independent users.	Political destabilisation with the aim of promoting the interests of another country

A comparative analysis shows that, despite differences in the tools of disinformation, their underlying mechanisms are identical. These include the escalation of uncertainty, a reduction in the transparency of the information environment, the transformation of perceptions of events, and the artificial promotion of targeted narratives. The manipulation of public opinion, along with the accompanying political polarisation and destabilisation of society, are key factors in the decline in trust in digital content.

The nature of information manipulation is becoming an increasingly prominent aspect of global politics. One of the key elements of the broader context is the ongoing global competition for leadership in the fields of artificial intelligence and machine learning. This primarily concerns the world's two leading technological hubs – the US and China – as well as the European Union and Russia. The ability to create disinformation is spreading rapidly; in the long term, its potential will depend on access to neural networks, machine learning, deep learning and artificial intelligence algorithms.

4. Discussion

In the second half of the 20th century, research was conducted into the possibilities of controlling human behaviour and into the informational, cognitive and psychological influences on individuals and society, which led to the transformation of theoretical principles into technologies. The most important characteristics of this new form of presenting information in a digital environment are its non-linearity, the multiplicity of virtual structures, its incompleteness and virtualisation (Хартунг&Брейдо, 1996). Numerous studies on disinformation and false information, conducted across a wide range of disciplines—including media studies, defence studies, and political science, have confirmed the focus of experts and the academic community on the multimodal nature of information dissemination and its consequences.

The 'Gutenberg era' was an era of predominantly monomodal communication. Kress and van Leeuwen drew attention to this, noting that the most significant genres of printed communication in Europe – academic books, novels, and official and political documents—were virtually devoid of illustrations; musicians wore identical uniforms and could not afford (with the exception of the conductor and, occasionally, soloists) to express emotions on stage, and so on. (Brandt, 2001). However, modern digital communication, by integrating spoken and written language, static images (photographs, graphics), moving images (video) and audio (music, sounds), has opened the door to a multimodal world encompassing visual, linguistic, auditory, spatial and gestural communication.

For example, the cross-platform messaging app 'Telegram' allows you to:

- Store an unlimited number of files and documents;
- Exchange text, voice and video messages;
- Create and manage microblogs;
- Develop and use bots;
- Use secret chats with two-factor authentication and multiple encryption protocols, ensuring maximum anonymity and privacy (Манёрова & Манёров, 2022).

In a multimodal environment, the accelerated escalation of the problem of disinformation and fake news has become virtually inevitable. In his concept of 'post-truth', the American sociologist Steve Fuller identifies two main factors contributing to the collapse of institutions that were once capable of verifying claims about the truth – institutions that once defined a 'more cohesive public sphere' – leaving an epistemological vacuum filled by citizens. (Fuller, 2018):

- The highest level of education in the history of the world's population,
- The availability of the internet and digital networks for disseminating information.

Digital competences are becoming a fundamental requirement in modern society. According to the European DigComp 2.2 framework, digital competence involves the confident, critical and responsible use of digital technologies, as well as interaction with them in the context of learning, work and participation in society, encompassing information and digital literacy, communication and collaboration, media literacy, the creation of digital content (including programming), safety (including digital wellbeing and cybersecurity-related competences), as well as issues relating to intellectual property, problem-solving and critical thinking (Arribas et al., 2025).

Media literacy programmes aim to develop critical thinking skills and teach people how to evaluate sources of information, making them less susceptible to fake news (Bergstrom et al., 2018). There is a positive correlation between educational attainment and awareness of fake news. People with a higher level of education tend to be more discerning when it comes to news and are better equipped to spot fake news (Pennycook & Rand, 2019).

Research shows that older generations are often more aware of fake news than younger generations. Young people, who are more immersed in digital media, may be exposed to greater volumes of fake news and disinformation (Pennycook & Rand, 2019). It would appear that resilience to disinformation and fake news arises when there is low social polarisation, authoritative state-affiliated media systems, a high level of media literacy and a high level of trust within society, as noted by Hamprecht, Balčitenė and Horowitz. People with more conservative ideological views, a lower level of education, and less trust in state institutions, who rely on intuitive rather than analytical thinking, are more likely to spread disinformation (Nan et al., 2022). A study by the Stanford History Education Group (2023) found that users with low digital literacy are three times more likely to share conspiracy theories (Reichert et al., 2023).

What tactics, tools and skills are available for identifying and countering disinformation and false information? International cooperation: UN initiatives, such as General Assembly Resolution A/RES/77/247 (2022), have created a global platform for countries to share experiences. In 2024, 89 states joined the Global Partnership on Artificial Intelligence, developing common standards for labelling synthetic content (OECD, 2026). However, differences in approach remain: whilst the EU relies on self-regulation by platforms, state control prevails in authoritarian regimes, which limits cross-cultural dialogue.

One of the most logical steps in combating misleading content is the development of user-driven fact-checking services and recommendations from various online resources. Modern fact-checking combines artificial intelligence, crowdsourcing and expert analysis. The FactCheck.org platform, which has been operating since 2003, uses NLP models to verify automatically politicians' statements, achieving 98 per cent accuracy. (Poynter Institute, 2024). Metadata analysis tools such as InVID make it possible to detect video manipulation: in 2023, they were used to expose 1,200 deepfake videos linked to the elections (EU DisinfoLab, 2024). The Poynter Institute has launched an important initiative to bring together fact-checkers from around the world—its International Fact-Checking Network not only organises conferences and monitors the formats and trends in the spread of fake news, but also provides online training. Recent advances in machine learning and natural language processing are opening up unprecedented opportunities for the large-scale detection of disinformation on the internet.

Several reviews provide an overview of current models for detecting fake information online:

- Aïmeur, Amri and Brassard analyse a multimodal method for detecting fake news based on image and text similarity, using computer vision (Aïmeur et al., 2023);
- Ozbay and Alatas propose a two-stage approach: the first stage involves extracting structured data from an unstructured news article, followed by the application of artificial intelligence algorithms in the second stage (Ozbay & Alatas, 2020);
- Zhang and Gorbani have developed a comprehensive three-tiered ecosystem for detecting fake news (the alert tier, the detection tier and the intervention tier). Machine learning and deep learning methods, utilising relevant datasets with carefully selected features, demonstrate particularly high accuracy rates in detecting falsified content (Zhang & Ghorbani, 2020).

In this regard, the following are seen as methods for combating digital disinformation:

- Legislative measures and regulation to combat digital disinformation;
- Removing social media platforms' legal immunity from content posted by their users;
- Technologies including content detection, authentication and the prevention of disinformation;

The investigation into disinformation on X (formerly Twitter) against the backdrop of the war between Israel and Hamas marked the first such action under the EU's new Digital Services Act. According to the final text of the legislation, tech giants operating 'very large online platforms' may be held liable for illegal and falsified content. In the event of serious breaches, a company could face a fine of up to 6 per cent of its total revenue. Mark Zuckerberg's Meta also received a warning from the European Commission. The company was urged to remove illegal content and "be extremely vigilant to ensure strict compliance with the rules".

5. Conclusion

Recommendation. Countering disinformation campaigns is a complex challenge that requires the development of systematic mechanisms for oversight and coordination at the international level. Implementing a multilateral approach involving government institutions, technology companies, civil society and international organisations has the potential to stabilise the information environment and restore the verifiability of digital content.

Feasibility moderate. The design of comprehensive systems to counter disinformation is characterised by a moderate degree of feasibility due to the asymmetry of interests among key actors. Whilst technological integration is technically feasible, the coordination of the regulatory framework between states with polarised foreign policy strategies faces institutional resistance. Multilateral cooperation between the civil sector, IT corporations and international institutions is feasible primarily within the framework of regional alliances or narrowly focused expert consortia, where the economic and defence costs for participants are minimised.

A compromise. Strengthening information security and purging the digital environment of destructive narratives automatically increase the risks of covert political censorship, restrictions on citizens' digital rights, and a reduction in online privacy. In order to develop effective, coordinated measures at the international level, sovereign states are forced to compromise and partially delegate their regulatory functions to supranational arbitrators or independent international organisations. However, against the backdrop of contemporary geopolitical polarisation, states' desire to protect their digital sovereignty and the autonomy

of their national segment of the internet directly hinders attempts to establish a unified global legal regime.

Assessment indicators. The number of permanent data-sharing channels between regulators, international organisations and IT giants. The proportion of independent civil society fact-checking platforms integrated into supranational monitoring mechanisms. Trends in public trust in digital communication channels, as indicated by cross-national sociological surveys.

Recommendation. Funding independent media organisations through diversified grant schemes and crowdfunding models serves as a means of reducing information asymmetry in the digital environment. This mechanism helps to scale up verified content and increase transparency in the media market. Given the cross-border nature of information threats, institutionalising efforts to counter disinformation requires the establishment of multilateral regulatory frameworks under the auspices of relevant international organisations.

Feasibility moderately low. Given the cross-border nature of information threats, institutionalising efforts to counter disinformation requires the establishment of multilateral regulatory frameworks under the auspices of relevant international organisations. The main obstacle to the development of uniform legal frameworks is the divergence of national interests among key geopolitical actors and differences in interpretations of the limits of freedom of expression.

A compromise. Global regulatory frameworks require states to transfer some of their sovereign regulatory functions to supranational institutions. Compliance with strict cross-border content verification protocols slows down the media's response to dynamic information attacks.

Evaluation indicators. The number of states that have incorporated standards for cross-border legal regimes into their national legislation. An increase in the number of citations and traffic volumes of independent journalistic organisations within the overall media landscape. Trends in audience trust in digital news sources, based on data from annual opinion polls.

Recommendation. Countering disinformation campaigns requires interdisciplinary approaches that integrate technological advances, media literacy education, accountability of IT platforms, and collaboration between the academic community, politicians and media organisations.

Feasibility medium-low. Ensuring the legal and corporate accountability of transnational IT platforms faces resistance from the technology sector, which defends the commercial secrecy of its algorithms. The feasibility of this component depends directly on the scale of the regulator: sovereign states have limited leverage over extraterritorial corporations, whilst large supranational bodies are capable of enforcing accountability protocols. In turn, the development and implementation of digital content detection methods, artificial intelligence algorithms for identifying coordinated inauthentic behaviour, and automated fact-checking are at an advanced stage of technical readiness.

A compromise. The integration of the scientific community with government bodies enables expert recommendations to be implemented swiftly, but risks compromising the independence of science and leading to researchers being used as tools. Cross-sectoral cooperation between all stakeholders (IT professionals, the media, academics and civil servants) ensures a balanced approach, but complicates bureaucratic coordination, thereby slowing down the response to rapidly evolving information crises. Educational projects in the field of media literacy build society's fundamental resilience to manipulation; however, they require significant budgetary investment and are unable to counter current, short-term threats.

Evaluation indicators. The number of active regulatory and educational projects developed at the intersection of the IT sector, academia and government. Changes in citizens' critical thinking scores based on standardised testing results before and after completing educational programmes. The percentage of recommendations from the academic community and regulators that have been successfully incorporated into social media moderation algorithms.

References

- Adversarial Threat Report (2026). [Online] <https://transparency.meta.com/sr/first-half-2026-Adversarial-threat-report>
- Aïmeur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1), 30. <https://doi.org/10.1007/s13278-023-01028-5>
- Arribas, C. M., Arcos, R., & Gertrudix, M. (2025). Rethinking education and training to counter AI-enhanced disinformation and information manipulations in Europe: A Delphi study. *Cogent Social Sciences*, 11(1), Article 2501759. <https://doi.org/10.1080/23311886.2025.2501759>
- Balkin, J. M. (2017). Free speech in the algorithmic society: Big data, private governance, and new school speech regulation. *UC Davis Law Review*, 51(3), 1149–1210. <https://doi.org/10.2139/ssrn.3038939>
- Bernays, E. L. (2013). *Public relations*. University of Oklahoma Press. https://books.google.az/books?hl=ru&lr=&id=wBFP_qrOYk8C&oi=fnd&pg=PP2&dq=a+s+1945,+Edward+Bernays,+in+his+book+%C2%ABPublic+Relations%C2%BB&ots=G2_99FVoZr&sig=QuWt7nsHwZhmLgxKWJp5ma4Tc4M&redir_esc=y#v=onepage&q=a+s%201945%2C%20Edward%20Bernays%2C%20in%20his%20book%20%C2%ABPublic%20Relations%C2%BB&f=false
- Brandt, S. (2004). Multimodal discourse: The modes and media of contemporary communication [Review of the book *Multimodal discourse: The modes and media of contemporary communication*, by G. Kress & T. van Leeuwen]. *Language in Society*, 33(1), 115–118. <https://doi.org/10.1017/S00474045042210544>. Digital 2025:
- Digital News Report 2024 // Reuters Institute for the Study of Journalism. (2024).[Online] <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024>)
- Eisele I., Steinwehr U. Fact check: AI fakes in Israel's war against Hamas. (2023) [Online] <https://www.dw.com/en/fact-check-ai-fakes-in-israels-war-against-hamas/a-67367744>
- Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, 37(2), 145–156. <https://doi.org/10.1080/10584609.2020.1723755>
- Fukuyama, F. (2021). The future of platform power: Solving for a moving target. *Journal of Democracy*, 32(3), 173-177. <https://doi.org/10.1353/jod.2021.0044>
- Fuller, S. (2018) Post- Truth. Knowledge as a Power Game. London-NY: Anthem Press. <https://doi.org/10.2307/j.ctvgd30v>
- Goswami, R. (2023, October 9). X, formerly Twitter, amplifies disinformation amid the Israel-Hamas conflict. CNBC. <https://www.cnbc.com/2023/10/09/x-formerly-twitter-amplifies-disinformation-amid-the-israel-hamas-conflict.html>
- Kuehl, D. T. (2002). “Information operations, information warfare, and computer network attack: Their relationship to national security in the information age”. In M. N. Schmitt &

- B. T. O'Donnell (Eds.), Computer network attack and international law (Vol. 76, pp. 35–58). Naval War College Press.[Online] <https://digital-commons.usnwc.edu/cgi/viewcontent.cgi?article=1400&context=ils>
- Mölder, H., & Sazonov, V. (2018). Information warfare as the Hobbesian concept of modern times—The principles, techniques, and tools of Russian information operations in the Donbass. *The Journal of Slavic Military Studies*, 31(3), 308–328. <https://doi.org/10.1080/13518046.2018.1487204>
- Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. *Physica A: statistical mechanics and its applications*, 540, 123174. <https://www.sciencedirect.com/science/article/abs/pii/S0378437119317546> [Online]
- 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*, 116(7), 2521–2526. <https://doi.org/10.1073/pnas.1806781116>
- Qorib, F., Oktarina, R. A., Mardiana, S., & Wulandari, E. R. (2024). Clickbait in the era of information technology: Evaluating engagement strategies and ethical journalism in online media. In *2024 Seventh International Conference on Vocational Education and Electrical Engineering (ICVEE)* (pp. 204–209). IEEE. <https://doi.org/10.1109/ICVEE63912.2024.10823912>
- Reichert F., Pan Q., Chen L. L. (2023) “Digital Literacy Assessment” (paper commissioned for UNESCO’s 2023 Global Education Monitoring Report, Technology in education) // Unesco. 65 p.[Online] <https://unesdoc.unesco.org/ark:/48223/pf0000386202>
- Steinebach, M., Bader, K., Rinsdorf, L., Krämer, N., & Roßnagel, A. (Eds.). (2020). *Desinformation aufdecken und bekämpfen: Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität* (Vol. 45). Nomos. <https://doi.org/10.5771/9783748904816>
- Suciu, P. (2022) “Is Russia’s Invasion of Ukraine The First Social Media War?” [Online] // Forbes. <https://www.forbes.com/sites/petersuciu/2022/03/01/is-russias-invasion-of-ukraine-the-first-social-media-war/>
- Tesich, S. (1992). A government of lies. *The nation*, 254(1), 12–15. <https://go.gale.com/ps/i.do?id=GALE%7CA11665982&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=00278378&sw=w&p=AONE&userGroupName=anon%7E4d324676&aty=open-web-entry>
- Zhang, X., & Ghorbani, A.A. (2020). An overview of online fake news: Characterization, detection, and discussion. *Inf. Process. Manag.*, 57, 102025. <https://doi.org/10.1016/j.ipm.2019.03.004>
- Манёрова К. В., Манёров Р. В. (2022) “Мультимодальный тип текста «сообщение» в мессенджере «Телеграм»” [Online] // (на материале немецкоязычного канала партии «Альтернатива для Германии») // *Немецкая филология в Санкт-Петербургском государственном университете*. Вып. 12. С. 273–287. <https://cyberleninka.ru/article/n/multimodalnyy-tip-teksta-soobschenie-v-messendzhere-telegram-na-materiale-nemetskoyazychnogo-kanala-partii-alternativa-dlya>
- Сулейман-заде Л.С. (2024) «Дипфейки как разновидность массовой цифровой дезинформации». Журнал «Geostrategiya», 1(79), стр.67-72 <https://doi.org/10.30546/3006-0346.2024.1.79.66>

- Сулейман-заде Л.С. (2025) «Политические алгоритмы в международной цифровой коммуникации». Журнал «*Geostrategiya*», 3(87), стр.89-94
<https://doi.org/10.30546/3006-0346.2025.3.87.529>
- Сытник, А. Н. (2022). Роль «цифровых публик» в информационной войне об Украине в мире. в Беларусь в современном мире: *Материалы XXI Международной научной конференции, посвященной 101-й годовщине образования Белорусского государственного университета*. Белорусский государственный университет. Беларусь, Минск [Online] (стр. 119-126).
<https://elib.bsu.by/bitstream/123456789/289659/1/119-126.pdf>
- Хартунг Ю., Брейдо Е. (1996) Гипертекст как объект лингвистического анализа. - В: "Вестник Московского университета", сер. 9, Филология, 1996, N3, с. 61-75.
<https://cyberleninka.ru/article/n/gipertekst-ponyatie-osnovnye-harakteristiki-vozmozhnye-podhody-k-lingvisticheskomu-analizu>