

Deepfakes, Fake News, And Digital Trust: Insights from Romanian Social Media Users

Mirela ABRUDAN¹ , Carmen Ioana NEAMȚU²

Abstract. In today's digital society, the spread of misinformation through false online content has become a major problem worldwide. Fake news, deceiving information presented as real, has the potential to manipulate public opinion, distort facts, and weaken trust in traditional and online media sources. With technological advances, a new form of misinformation has emerged known as deepfakes, which makes distinguishing between reality and fiction even more challenging. Deepfakes are a more complex form of digital manipulation, using artificial intelligence to generate apparently genuine but in fact completely fake audiovisual content. The present study analyses the impact of fake news and deepfakes on Romanian social media users by conducting a non-representative self-administered online survey (N=154). The results indicate that the effects of fake news and deepfakes are influenced by socio-demographic factors such as age, gender, and educational level. Young people, as well as Romanian social media users with a high level of education, are more familiar and aware of these phenomena and their negative effects, likely due to higher media literacy. The collected data also suggest that female users have a critical and cautious approach regarding the consumption of digital content, thereby reducing the impact of misinformation. Romanians also believe that repeated exposure to fake news and deepfakes

¹ Assoc. Prof. Ph.D., Department of Communication, PR, and Advertising, Babeș-Bolyai University, mirela.abrudan@fspac.ro

² MA Communication and Advertising, National School of Political and Administrative Studies SNSPA, ioananeamtu41@yahoo.com.



leads to a loss of trust in social media. However, most users do not consider that being repeatedly confronted with fake news and deepfakes makes them necessarily more susceptible to manipulation.

Keywords: fake news, deepfakes, misinformation, disinformation, manipulation, social media, Romania.

Introduction

In recent years, communication technologies and artificial intelligence have amplified the propagation of fake news and deepfakes in the digital environment. Due to the outbreak of the COVID-19 pandemic, the war started by Russia in Ukraine, but most of all the electoral campaign, followed by the annulment of the first round of the 2024 presidential elections, disinformation and manipulation of public opinion also reached alarming levels in Romania (Stănescu, 2024). Thus, the spread of false content poses a major challenge to the integrity of public discourse and the stability of democratic societies. Fake news, deceiving information presented as real, uses social media platforms to spread rapidly and often manipulates public opinion for political, financial, or ideological purposes. Deepfakes, a more advanced form of technology using artificial intelligence to produce fake video or audio content, further complicate the information landscape by undermining trust in visual and auditory evidence. Taken together, these forms of disinformation undermine public trust in media institutions and democratic processes, highlighting the urgent need for robust detection and awareness mechanisms to mitigate their impact (Botha & Pieterse, 2020). In October 2022, the EU adopted the Digital Services Act, which should have been fully effective by February 2024, a regulation applied to online intermediaries, including social networks, meant also to tackle disinformation. To do so, online platforms must hire fact-checkers, maintain content moderation teams across all official EU languages, and implement rapid alert systems, especially around elections. Content generated by artificial intelligence, such as deepfakes, must also be marked correspondingly (O'Carroll, 2024).

Numerous studies and reports have explored how fake news and deepfakes affect Romanian social media users. The most analysed topics in the context of digital communication and fake content rank among exposure and

vulnerability perceptions; engagement on online platforms generated by fake content; fake news, algorithms, and political influence; deepfake and cyberbullying; deepfake scams, fraud and propaganda, or election manipulation.

A national survey conducted by INSCOP Research at the request of the press agency News.ro (2024) found that over 70% of Romanians believe they are exposed to fake news, and 43% say social networks present the highest perceived risk for disinformation. Moreover, 75.9% of the Romanian citizens think that their voting choices are affected by disinformation and fake news (INSCOP, 2025). Reuters Institute Digital News Report from 2024 emphasises Romanians' reliance on online platforms when it comes to news consumption, and the rising concern about misinformation (Radu, 2024). The high occurrence of manipulation mechanisms on Romanian social media during the 2024 elections was marked as an unprecedented phenomenon in the country's history, resulting in the annulment of the first round of the presidential election, due to the electoral process being deemed as flawed. Online disinformation distorted public opinion and agenda-setting (Stănescu, 2024). TikTok was used for at-scale manipulation attempts to reach Romanian users (Global Witness, 2024). Network analysis shows bot-driven amplification on anti-EU / NATO narratives and cites around 34% fake-profile engagement on key posts, illustrating how coordinated inauthentic behaviour can shape perceived sentiment among social media audiences (Baruchin, 2025). Bitdefender Labs reported numerous scam lures during 2024, from fake celebrity endorsements to AI audio and video deepfakes, cloned media brands and fabricated financial testimonials used in social media ads, advising Romanians how to manage such practices (Bizgă, 2025). The 2025 Digital News Report highlights audiences' growing concern about synthetic content and misinformation, reinforcing why Romanian social media users report vigilance and avoidance (Newman, 2025).

As the spread of fake news and deepfakes continues to pose significant challenges to social trust, media integrity, and democratic processes, individuals, institutions, and governments must urgently develop critical thinking skills, media literacy initiatives, and technological solutions to combat the propagation of falsehoods and defend the reliability of public discourse.

On the 1st of April 2024, the Romanian National Cybersecurity Directorate released a guide designed to help identify deepfake content. The document provides useful information on the importance of understanding the phenomenon, technologies behind it, practical examples and warning signs,

detection tools and methods, the role of education in countering deepfakes, along with useful awareness tips, and recommendations in case of becoming a deepfake victim (JURIDICE.ro, 2024).

In this context, the current study surveys Romanian social media users to map their awareness and perceptions of negative influences of fake news and deepfakes in the digital environment.

Fake news

Although the concept of fake news is often associated with the digital age and the emergence of online platforms, the spread of misinformation and propaganda precedes the Internet by many centuries (Winston & Winston, 2020). From ancient times, misinformation, disinformation, and propaganda have shaped public perception, influenced political discourse, and had a profound impact on social dynamics (Soll, 2016). The principle of fake news can be traced back to ancient civilisations, where several means of propaganda were used by rulers and elites to maintain power and control over the masses. Political leaders and religious authorities resorted throughout history to censorship, manipulation of historical narratives, and the dissemination of fabricated stories to strengthen their authority and suppress dissenting opinions (Soll, 2016).

Over the past decade, the concept of fake news has become a prominent and controversial topic on the media scene, as well as in politics and social discourse. Coined in the late 2000s and propagated during the turbulent political events of the 2016 presidential election in the U.S., fake news has evolved from a niche term to a ubiquitous and multifaceted phenomenon with far-reaching implications for public trust, the integrity of information, and democratic processes (Allcott & Gentzkow, 2017).

The propagation of misleading narratives and conspiracy theories across social media platforms, along with concerns about foreign interference and algorithmic amplification, influenced public opinion and generated a global debate about the spread and impact of misinformation in the digital age (Allcott & Gentzkow, 2017; Tandoc et al., 2018). In response, academics, journalists, policymakers, and technology companies have joined their efforts to limit and combat the dissemination of fake news through interdisciplinary research, fact-checking initiatives, media literacy programs, and regulatory intervention (Tandoc et al., 2018).

Despite these efforts, the phenomenon remains a challenge today, fuelled by partisan polarisation, echo chambers, and information silos, extending beyond politics to public health, climate change, social cohesion, and presenting complex ethical, legal, and technological challenges for actors in various fields (Gelfert, 2018). The study of fake news represents, therefore, a dynamic and interdisciplinary area of research, reflecting the urgent need to develop mechanisms for combating misinformation in an increasingly connected and digital world.

Fake news encompasses not only inaccuracies but also content deliberately created to manipulate and influence journalistic discourse (Rini, 2017). It may contain partially incorrect information due to journalistic negligence, whether because of the lack of experience, or failure to check sources. However, in such cases, journalists do not intend to mislead their audience or hide the truth (Baptista & Gradim, 2022). Although most scholars support the theory that fake news is not entirely fabricated stories, some voices support this view (Buckingham, 2019). Fake news uses sensational, provocative, and controversial narratives that aim to exploit social stereotypes and prejudices, thus producing strong emotional responses (Rochlin, 2017). This approach facilitates the virality of false information, as such content is more likely to be shared. The profitability of fake news depends on its popularity, with Rini (2017) arguing that it is a form of mass manipulation and explaining that the degree of untruthfulness is vital for its perpetuation, whether for ideological or pecuniary reasons.

The phenomenon of fake news is a complex and multifaceted challenge with profound implications for society, democracy, and information integrity. It represents more than just misinformation, involving the deliberate manipulation of facts, the dissemination of false representations, and the exploitation of societal vulnerabilities for various purposes, whether political, ideological, or financial (Gelfert, 2018). Furthermore, the dissemination of fake news in the digital age underscores the need for robust strategies to combat misinformation, including initiatives to improve media literacy, fact-checking mechanisms, and ethical journalistic practices. Critical thinking and digital competence are important for the public to distinguish facts from fiction in an increasingly complex media landscape. Tackling the problem of fake news requires joint efforts by various actors, including governments, technology companies, media organisations, and civil society, to promote a better-informed, more resilient, and more democratic society (Tandoc et al., 2018).

Fake news on social media

In today's digital world, the propagation of fake news via social networking sites is a complex and multifaceted process. As individuals navigate a flood of information, the dissemination of misleading or fabricated content can cause confusion, disruption, and undermine trust in societal institutions (Arin et al., 2023). The high occurrence of false information on online platforms has concrete effects on individuals' beliefs and behaviour, and even on democratic processes. Creators and providers of such content can exploit vulnerabilities in social media algorithms to manipulate public discourse for ideological, political, or financial purposes (Arin et al., 2023).

The accessibility and affordability of creating social media profiles also contribute to the proliferation of harmful actors, like social bots, cyborg users, and trolls (Shu et al., 2017). Social bots, automated accounts controlled by computer algorithms, work autonomously to create content and interact with human users or other bots on social networks (Ferrara et al., 2016; Shu et al., 2017). These entities can play a harmful role by deliberately spreading misinformation and manipulating online discourse, as demonstrated by their widespread influence during the electoral campaign before the 2016 U.S. presidential elections. The decentralised nature of Twitter, officially known as X since 2023, and its rapid transmission capabilities, make it vulnerable to the spread of misinformation and disinformation (Drolsbach & Pröllochs, 2023). Fake news also has the potential to destabilise the integrity of the news ecosystem by eroding its foundations of authenticity and trust. For example, during the 2016 U.S. presidential elections, it became clear that the most popular deceiving stories on Facebook were shared more than the most popular authentic news in the mainstream media (Allcott & Gentzkow, 2017).

At the same time, trolls, real human users who are experts at causing chaos, disrupting online communities and provoking emotional reactions from users, also exert a significant influence on the propagation of fake news in the digital realm. Their deliberate intention to arouse negative emotions such as anger and fear, which in turn cause doubt, mistrust, and irrational behaviour among users (Shu et al., 2017), exacerbates the dissemination of disinformation in digital spaces and highlights the complex interplay between human behaviour and technological platforms in today's information ecosystems (Cheng et al., 2017). According to reports, there were around 1,000 paid Russian trolls involved in spreading fake news about Hillary Clinton. The behaviour

of trolls is particularly influenced by the emotional state of users and the contextual dynamics of online conversations, which facilitates the smooth spread of false stories in seemingly “natural” digital communities (Shu et al., 2017).

Cyborg accounts are created by humans as a cover, with automated programs performing actions on online platforms (Ferrara et al., 2016). The fluid transferability between human and automated functionality offers cyborg users various opportunities to spread disinformation. Overall, these actively partisan and productive malicious accounts exert considerable influence as powerful and amplifying sources of fake news in social media ecosystems (Shu et al., 2017).

With the growing importance of social networks, people have become increasingly reliant on these digital channels to keep up to date with current events. However, a major problem permeating these news feeds is the ubiquitous presence of false or misleading content. Despite the efforts of some online platforms to curb the propagation of fake news through features such as tagging or warnings, the persistence of false information has profound and lasting effects. Fake news not only misleads people but also creates a climate of uncertainty and mistrust that undermines the very foundations of informed public discourse. Furthermore, its uncontrolled spread can fuel social unrest, exacerbate existing tensions, and even hinder democratic processes by undermining the integrity of public discourse and decision-making in elections (Arin et al., 2023).

Deepfake

The emergence and propagation of deepfakes have generated both excitement and concern in the fields of technology, media, politics and society in general. Deepfakes are a sophisticated form of synthetic media produced through artificial intelligence-based algorithms, whereby existing images or videos are digitally manipulated or overlaid in such a way that highly convincing but completely fake content is created. This innovative technology blurs the lines between facts and fiction and has profound ethical, social, and political implications (Narayan et al., 2022). While deepfakes offer new opportunities for creative expression and entertainment, their potential for abuse and exploitation raises critical questions about misinformation, privacy

violations, and the loss of trust in both visual and audio media. Therefore, understanding the complexity and implications of this phenomenon is important for policymakers, media professionals, and the public (Chesney & Citron, 2019).

Deepfakes convincingly imitate human behaviour and appearance, extending from visual environments to audio manipulation. Technological progress has made deepfakes almost indistinguishable from real images, presenting a range of challenges and opportunities for people around the world. The term “deepfake” originates from the fusion of “deep learning” and “fake” and illustrates the dependence on artificial neural networks in machine learning (Mirsky & Lee, 2021). Primarily used to manipulate and generate human images, this technology also has a variety of creative and useful functions. It can be found in numerous sectors, such as the fashion industry and e-commerce, where it simplifies consumer transactions. In addition, the entertainment industry uses deepfake technology to synthesise artificial voices for musicians and enhance film productions with sophisticated special effects (Chesney & Citron, 2019). Its therapeutic potential is also evident, considering the ability to enable simulated interactions for Alzheimer patients, potentially helping to preserve memory (Mankoo, 2023). In addition, various online communities have emerged that explore the comedic potential of deepfake memes, such as the creation of music videos.

Despite the positive outlook, however, deepfake technology has also become notorious for its unethical and harmful aspects. In November 2017, the first widely known case of deepfakes appeared, when a Reddit user known as “Deepfakes” used deep learning algorithms to insert the faces of celebrities into pornographic videos, sparking a media frenzy and an increase in the creation of similar content (Mirsky & Lee, 2021). Subsequent cases, such as the spread of a deepfake video file showing former U.S. president Barack Obama, have intensified fears about identity theft and the propagation of falsehoods on online platforms. Starting with these incidents, academic interest in deepfakes has grown exponentially (Barari et al., 2021), due to the significant ethical dilemmas and societal risks that the phenomenon presents. Prominent figures such as politicians and celebrities are particularly vulnerable to the creation of defamatory or pornographic deepfake content, which exploits the publicly available media in which they are portrayed. This ethical issue is exacerbated by the ease with which individuals can create misleading content, which can lead to malicious abuse, including the dissemination of

unauthorised or fake images and audio files. Consequently, the spread of untraceable artificial content poses a major challenge to authenticity, integrity, and privacy in the digital environment (Akhtar, 2023).

Deepfake on social media

Even if technological advancement has generated new possibilities for creative expression, entertainment, and educational purposes, the propagation of deepfakes on social networking sites poses formidable challenges, including the potential for widespread misrepresentation, manipulation of public perception, and undermining trust in digital media (Chesney & Citron, 2019). Furthermore, deepfakes raise profound ethical and societal questions, such as privacy violations, identity theft, and confusion between truth and fiction. Given the evolution and propagation of deepfake technology, researchers are working to comprehensively understand its multifaceted repercussions on the dynamics of social media, the integrity of information, and broader societal discourse. Through rigorous research and analysis, scientists are attempting to shed light on the complexity of deepfake technology and develop strategies that aim to diminish its negative consequences while harnessing its capacity for positive societal impact (Akhtar, 2023).

The dissemination of digital photos and videos has increased significantly due to the widespread use of smartphones and the growing popularity of online platforms. This has led to social networks becoming one of the most important sources of daily information and news consumption. Among these platforms, X, formerly Twitter, has a decentralised and dynamic nature and stands out with a significant active user base, particularly as a fast news disseminator compared to traditional media (Fagni et al., 2021). However, this also leads to an increase in spoofing techniques aimed at quickly spreading false information. The fake content generated is carefully designed to appear authentic and often focuses on current topics to accelerate its spread and involve more people in false stories, ultimately posing significant risks to society (Tolosana et al., 2020). Such multimedia content can extend widely in an uncontrolled manner and creates an alternative narrative of reality, while identifying the originators of rumours remains a major challenge due to the difficulty of determining the source of the published media material (Narayan et al., 2022).

Also, on Facebook, one of the world's most popular social media channels, there is a serious risk that deepfakes could be used to propagate inaccurate information and distortions. This development not only calls into question the credibility of content but can also have a significant impact on user trust in the platform. Given these challenges, Facebook must develop effective strategies and technologies to detect and combat deepfakes while strengthening the awareness and digital media literacy of its users (Shu et al., 2017).

Fake content poses a significant digital threat. In times of heightened sensitivity or during major events, the propagation of falsehoods becomes more frequent. During elections, for example, there is an increased risk that opponents would exchange deepfake videos to undermine each other. Although deepfakes can have a profound impact on society, society itself plays a crucial role in combating this type of content (Gunther et al., 2018). The continued dissemination of fake videos on the Internet could undermine trust in online information for future generations. Society must therefore recognise that sharing deepfakes inadvertently helps their creators to spread misinformation. The social impact of deepfakes technology is profound and could lead to people being reluctant to distribute personal content on social networking sites due to trust issues, which could have far-reaching consequences for all industries (Chesney & Citron, 2019).

Given their immense influence, online platforms such as Facebook, Instagram, TikTok, or X have been viewed critically for their potential to quickly spread inappropriate content worldwide. Social networks have been used to promote terrorist violence, distort election results, and influence public perception of important democratic events.

Although working to implement strategies and technologies to combat fake news and deepfakes, the key challenge remains to ensure the authenticity of content and strengthen user trust in these platforms (Fagni et al., 2021). A comprehensive approach is needed, based on technological innovation, community collaboration, improved digital media literacy, and policies to regulate user engagement, to minimise the impact of disinformation in the digital realm.

Research design

This research aims to identify the perceived effects that fake news and deepfakes occurring on online platforms might have on Romanian social media users from different age cohorts, gender, and educational backgrounds. To investigate whether individuals are aware of these phenomena and their possible negative impact, as well as the strategies they use to avoid misinformation and manipulation in the digital environment, a survey was designed and conducted.

Based on the literature at the intersection of social media use, effects of this behaviour, and fake news and deepfakes, a series of research questions and hypotheses were established, to be addressed via the quantitative research method employed:

RQ1: Do socio-demographic criteria influence the recognition potential and the vigilance of Romanian social media users towards fake online content?

H1: The awareness and vigilance regarding fake news and deepfakes in the digital landscape are influenced by factors such as age, gender and, educational level.

RQ2: How are Romanian social media users affected by the exposure to fake news and deepfakes on social media?

H2: Users who are repeatedly exposed to fake news and deepfakes on social media are more susceptible to misinformation and manipulation.

H3: Frequent exposure to fake news and deepfakes on social media leads to a loss of public trust in these platforms and increases concern towards the consumption of manipulative online content.

Results

Through empirical analysis and critical examination, the current research intends to analyse the impact of fake news and deepfakes on Romanian social media users and thus, to contribute to a deeper understanding of the complex dynamics shaping today's digital communication and the consequences for individuals, communities, and democratic societies in general. Ultimately, the study aims to provide insights that can inform strategies to diminish the negative effects of fake news and deepfakes, to promote an informed and resilient information landscape in the digital age.

To address the research questions and hypotheses above, a quantitative research method was employed. A Romanian-language survey was developed and disseminated through Facebook and Instagram over the period April – May 2024. The questionnaire gathered N=154 valid responses. The sample was a non-probability one, constituted through self-selection, and the survey started with a section for participants' informed consent.

Socio-demographic data of the participants play an important role in the present research, given the formulated research questions and hypotheses. Regarding the age segmentation of the respondents, 3.25% fell into the age category of teenagers under 18. Young adults (18-24 years old) embody the most strongly represented age cohort in the sample (34.42%). Due to their high digital literacy and online media consumption habits, they probably easily accessed the questionnaire. 16.23% of the sample was represented by individuals aged 25 to 34, participants between 35 and 44 years old gathered 27.92%, while the age cohort between 45 and 60 also represented 16.23% of the total number of respondents. Finally, individuals over 60 years old are the least represented (1.95%), probably due to restricted access to the Internet or social media networks, a lower media literacy or interest in such online activities. They are also the only age category with no respondents aware of the concept of fake news or deepfakes. Young generations, under 18 to 44 years old, are instead familiar with the phenomena (93%), and indicate more vigilance when accessing information, especially on social networks, due to the emergence and spread of such content (87%).

In terms of gender representation, the collected data reveal a significant discrepancy, with most of the participants being females (66.2%). The male gender was less represented, with 52 respondents (33.8%). There were no registered responses of individuals declaring themselves non-binary. This strong presence of women could be due to their increased interest in the topic. The results also imply that women indicate a greater ability to recognise and reject false or inaccurate content (84%) than men (72%). A plausible hypothesis could be that females are more attentive to the phenomena of fake news and deepfakes, being more concerned about the accuracy and authenticity of online information.

Concerning the educational level, 29.2% of the survey participants declared pre-university studies as the last or current form of education attended. The majority have a bachelor's degree (32.5%) and a master's degree (24.7%), 8.4% of the respondents indicate having also postgraduate studies, while 5.2%

avoided answering this question. These numbers, corroborated with the data regarding the recognition potential and vigilance towards fake news and deepfakes in the online environment, could indicate a possible link between educational background and awareness of or interest in misinformation and manipulation in social media. Participants with higher educational levels, bachelor's or master's degree and post-university studies, are more familiar and recognise fake news and deepfakes on online platforms more easily (92%) than participants with pre-university studies (61%); this might also be because they are probably more familiar with the concepts of critical thinking and information evaluation. However, it is important to emphasise that this analysis provides only preliminary evidence and that we cannot draw definitive conclusions about the causal relationship between age, gender, and educational attainment and attitudes toward misinformation. It is plausible that other individual factors, such as personal experiences or sociocultural background, play an important role in influencing these attitudes.

The analysis of the frequency of social media use reveals a clear prevalence of users who access social media intensively and daily (76%). 15.6% of all respondents reported using social media daily, but not intensively. Only 8.4% of the participants use online social media platforms occasionally. However, the level of trust in content consumed on social networks is rather low. The data indicates a small proportion of the sample (1.9%) trusting the information on social media to a very high degree, and 11.7% to a high degree. 81.2% mentioned partial trust, and 4.5% of the respondents have no trust at all in the information posted on online social networking platforms. Collected data also shows that users are generally concerned about the truthfulness and reliability of online content (91,6%), and the analysis indicated that intensive use of social media is doubled by a stronger desire to be aware of the existence of misinformation and to understand how it can influence perceptions and decision-making (83,1%).

Regarding the familiarity with the concept of fake news, the survey participants (97.4%) are aware of and generally recognise this phenomenon and its negative effects in the online environment. A very small percentage (2.6%) stated that they are not familiar with this specific term. However, this does not necessarily imply that they are unaware of the phenomenon of misinformation itself. When asked about the frequency of encountering fake news on social media, 64.9% of respondents are confronted daily with this phenomenon, while 24.7% come across fake news several times a week.

Only 4.5% of the questioned social media users say they rarely encounter fake news on social networks. Knowledge of the concept of fake news can lead to a more vigilant attitude and a better ability to recognise and evaluate false or manipulative information on social media platforms. This correlation suggests that education about misinformation can play an important role in combating this phenomenon and promoting a more responsible and well-informed online experience.

Switching to deepfakes, 83.1% of the respondents stated they know the meaning of the concept and 54.5% consciously recognised such content on social media. This high level of awareness can be attributed to the growing interest in this technology among the media and the public. On the other hand, 11% of the survey participants are unaware of the existence of deepfakes, and 37% mentioned they have not consciously recognised a deepfake on social networking platforms so far. Finally, 5.8% of the respondents preferred not to answer the question referring to the familiarity of the concept of deepfakes, and 8.4% avoided indicating their recognition potential of such content. These findings underscore the need for a critical approach and adequate public education in detecting and dealing with digitally faked content and highlight the complexity and diversity of techniques used in the production of deepfakes.

Asked about how they deal with the discovery of fake digital content, the most common reaction of the survey participants was that they report the video to the platform where it was found (37.7%). This suggests that online platforms should be responsible for removing fake content and protecting users from its potential negative consequences. In addition, 29.9% of the respondents declared being cautious and checking the information using multiple sources before taking further action. This approach reflects a desire to confirm the authenticity of online content and avoid the dissemination of false information over the Internet. Data analysis indicates that the sources survey participants use and trust most for the validation process of information are official government sources (55.2%), followed by independent news sources (34.4%), mainstream media (31.8%) and personal relationships, like family and friends (24%). Online social networks are perceived as the least trustworthy source of information (14.3%). However, there is also a notable percentage (13.9%) of individuals in the sample who mentioned sharing and commenting on fake news and deepfakes immediately. Such a reaction could indicate a tendency to distribute information without first checking

its authenticity, which could lead to the unintentional rapid spread of fake content throughout the digital environment. Finally, 14.9% of the respondents simply ignore the discovery of fake news or deepfakes. The reasons for this kind of reaction could be manifold, including a lack of interest in or awareness of the impact and importance of the authenticity of online content. Overall, these findings emphasise once more the importance of education and public awareness about how to deal with and combat the propagation of fake news and deepfakes on the Internet.

Regarding the impact of false online content, 48.1% of the survey participants do not consider that the occurrence and the intensive exposure to fake news and deepfakes on social networks affect them in any way. This may reflect a certain confidence in their ability to evaluate and filter information, on the one hand. However, survey data indicate that only 9.7% of individuals questioned found it very easy to make the distinction between fake content and real content online. For almost half of the respondents (49.4%), the distinction is perceived as easy, but they acknowledge that it requires conscious effort and attention. On the contrary, around 32.5% of the participants find it difficult to identify the truthfulness of online information, while 3.2% consider this distinction to be very difficult. On the other hand, one explanation for the results concerning the effects of fake news and deepfakes on Romanian social media users can rely on the “third-person effect” (Lev-On, 2017), as most respondents (83.1%) declared themselves fully aware of the negative impact of misinformation on the society and find it very important to combat this phenomenon through information and education. Despite this statement, most individuals (48.1%) who expressed themselves regarding the influence on their own person denied registering any negative effects. On the contrary, 38.3% of the respondents admitted that frequent exposure to fake news and deepfakes makes them more susceptible to misinformation and manipulation. This significant percentage indicates that in Romanian online communities, there exists awareness of the potential negative effects of false information and digital manipulation on individuals’ thinking and behaviour. Fake digital content can also undermine trust in truthful information and distort the perception of reality. There is also a rather high number of individuals in the sample (13.6%) who avoided expressing their opinion concerning this important aspect of our survey.

The necessity of social media platforms taking measures to detect fake news and deepfakes is considered by 81.8% of the survey participants

as very important. In addition, 16.9% of the respondents recognised this issue as being important. There were no responses that described this subject as less important or not important at all. These results show a broad consensus among respondents on a topic intensely debated in the literature, namely the importance of preventive measures and the implementation of effective strategies and technologies for managing and filtering manipulative content on the Internet (Fagni et al., 2021).

Automatic detection technologies are considered an important solution for limiting the dissemination of misinformation on social networking sites by most respondents (67.5%). Almost half of the individuals questioned (46.1%) mentioned the importance of greater transparency regarding the functioning of online platforms and their algorithms for recommending and filtering content. Educating users to recognise and handle false information is also strongly supported by 35.1% of the survey participants, while strict sanctions for misinformation are considered important by 50% of them. Overall, the responses indicate that combating misinformation on social networks requires a combination of advanced technological solutions, transparency on the part of social media platforms, education and awareness-raising campaigns among users, and the introduction of stricter penalties for those who propagate false information.

The answers concerning the responsibility for the spread of fake news and deepfakes on the Internet reveal different views. Most respondents (71.4%) hold social media platforms accountable for the propagation of false content, while 51.3% believe that individuals and companies bear a large responsibility for the spread of misinformation in the digital environment. As for the government, 35.7% of the survey participants believe that it should bear some responsibility for combating the dissemination of misinformation on social networking platforms. This view reflects a desire for government intervention and regulation in online communication to protect users and promote a safer and more reliable digital environment.

Important negative effects of the analysed phenomena are also growing concerns and a significant loss of trust in the online environment due to the spread of false content. 19.5% of the respondents stated that their trust in social media has decreased significantly, and their concern regarding manipulation amplified due to frequent emergence and exposure to fake news and deepfakes, 31.2% lost trust to a great extent, while 39.6% of the individuals questioned declared that their trust in social networks has

decreased to some extent. These findings suggest that the credibility of online platforms is rather low among Romanian social media users and urge for rapid development of strategies and measures to reduce the negative effects of misinformation, manipulation, and propaganda on the Internet.

The last question deals with measures that could help restore trust in information on social media, and the responses correspond to the results obtained regarding the solutions for limiting the spread of misinformation and digital manipulation, which survey participants mentioned. The collected data indicate that, according to Romanian social media users, stricter verification of information (85.7%), user education (47.4%), and greater transparency of algorithms (29.2%) are the most important ways to restore people's trust in online platforms. These findings emphasise the urgent need for developing strategies and initiatives to improve the quality and reliability of online content, as well as an educational approach focused on raising awareness of the characteristics of false information and how to deal with it.

Discussion and conclusions

The present research aimed to investigate the awareness and impact of fake news and deepfakes on Romanian social media users in the digital environment. If socio-demographic criteria influence the recognition potential and vigilance towards false and manipulative online content, or how intensive exposure to such content makes Romanian users more susceptible to misinformation and affects their trust in social networking sites, are some of the questions our endeavour tried to answer through an online self-administered survey. Despite the reliance on a convenience sample of limited size, which precludes a conclusive assessment of the hypotheses, the study nevertheless generates results of sufficient relevance to merit further discussion. Regarding RQ1 and the related hypothesis, the findings indicated that indeed variables like age, gender and educational background affect the awareness towards fake news and deepfakes in the digital environment. Thus, young generations, who grew up with technological progress and intensive use of social media platforms, have better skills to recognise these phenomena. Digital natives are generally more sceptical and vigilant about the information they consume online and use various methods and digital tools to verify its authenticity. Furthermore, Romanian female social media users seem to be aware of the risks associated with digital misinformation

to a greater extent than men, which prompts them to be cautious and reluctant to accept or share suspicious content. The level of education also plays an important role concerning awareness and understanding of fake news and deepfakes on the Internet. People with higher education are better informed, more familiar with the concepts, due to their higher media and digital literacy, and more likely to be able to recognise disinformation and complex digital manipulation.

H2 concerning the higher susceptibility to misinformation and manipulation in case of frequent exposure to fake news and deepfakes on social networks could not be confirmed, as most of the survey participants denied such influence. In this case, one of the factors that could have affected the validity of the results obtained is the phenomenon known as the “third-person effect hypothesis” (Bawazir & Idid, 2024). According to this, individuals tend to believe that they are less vulnerable to the influence of media messages than other people, and implicitly to an underestimation of the impact of fake news and deepfakes on oneself, while others are perceived as more susceptible. Thus, the questionnaire results are contradictory, because initially, most of the respondents acknowledged mistrust in social media, fuelled by frequent misinformation and sophisticated digital manipulation that undermines the credibility of information available online. Therefore, the collected data may not accurately reflect the actual perceptions and behaviours of users. This limitation could affect our understanding of the actual impact of disinformation phenomena in the Romanian digital landscape. However, acknowledging this may pave the way for further, more targeted, and longer-term research that more accurately examines the psychological effects of fake news and deepfakes on Internet users.

Romanian social media users are increasingly aware of the advanced techniques used to create fake news and especially deepfakes, which contribute to a widespread sense of distrust and caution. This rather general scepticism indicates that there is an urgent need for media education and the development of effective mechanisms for verifying information to restore trust in social media. Educational institutions, social media platforms, and regulatory authorities must work together to develop and implement media and digital literacy programs that equip users with the ability to recognise and combat misinformation. The results of the questionnaire analysis confirm, therefore, H3, that indeed, intensive exposure to fake news and deepfakes on social networking sites leads to a loss of public trust in these platforms and increases concern towards the consumption of manipulative online content.

As any such academic undertaking, the current study is inevitably constrained by certain methodological limitations and potential sources of error; nonetheless, it also elucidates promising directions for future research. Although the survey relied on a non-probabilistic, non-representative sample, the findings provide valuable insights that allow for a preliminary delineation of Romanian social media users' perception of fake news and deepfakes that occur daily in today's digital society. However, it must be mentioned that our study lacks a detailed analysis of the ethical perspective of the discussed topic, which is of great importance to understand the complexity of such phenomena. An ethical assessment would contribute significantly to clarifying the moral implications of the spread of disinformation and provide a solid basis for formulating counterstrategies that consider the principles of integrity and social responsibility. The advancement of deepfake technology can influence our perception of history and reality in general and poses major challenges to the understanding and interpretation of the past and present. The ability to create fake audio and video content with increasing accuracy raises fundamental questions about the authenticity and reliability of the information we consume, questions which are crucial for future research and critical engagement with the impact of deepfake technology in our society. Therefore, the ethical implications of the phenomena of fake news and deepfakes are profound and multifaceted, ranging from eroding public trust, manipulating political outcomes, and violating individuals' dignity and consent, to exposing regulatory inadequacies, endangering legal processes, and challenging norms around creativity and media literacy. Each dimension demands thoughtful solutions: legal frameworks, platform accountability, public education, and an ethical culture that balances innovation with responsibility.

References

- Akhtar, Z. (2023). Deepfakes generation and detection: A short survey. *Journal of Imaging*, 9(1), 18. <https://doi.org/10.3390/jimaging9010018>
- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211-236. DOI: 10.1257/jep.31.2.211
- Arin, K. P., Mazrekaj, D., & Thum, M. (2023). Ability of detecting and willingness to share fake news. *Scientific Reports*, 13(1), 7298. <https://doi.org/10.1038/s41598-023-34402-6>

- Baptista, J. P., & Gradim, A. (2022). A working definition of fake news. *Encyclopedia*, 2(1), 632–645. <https://doi.org/10.3390/encyclopedia2010043>
- Barari, S., Lucas, C., & Munger, K. (2021). Political Deepfake Videos Misinform the Public, But No More than Other Fake Media and (Sometimes) Real Media. *OSF Preprints*, 13, 1-16. <https://osf.io/cdfh3/>
- Baruchin, R. (2025, April 22). Analyzing Fake Profiles' Impact on Romania's Election Crisis. *Cyabra.com*. Retrieved from <https://cyabra.com/blog/analyzing-fake-profiles-impact-on-romania-s-election-crisis/>
- Bawazir, A. & Idid S. A. (2024). The impact of third-person effect on social networking sites privacy risks and protective measures adoption among Yemeni students in Malaysia. *Heliyon*, 10(9). DOI: 10.1016/j.heliyon.2024.e30386
- Bîzgă, A. (2025, May 16). Deepfake Scam Exploits Romanian Presidential Candidates to Lure Victims into Fake “Neptun Deep” Investment. *Bitdefender.com*. Retrieved from <https://www.bitdefender.com/en-us/blog/hotforsecurity/deepfake-scam-exploits-romanian-presidential-candidates-to-lure-victims-into-fake-neptun-deep-investment>
- Botha, J., & Pieterse, H. (2020). Fake News and Deepfakes: A Dangerous Threat for 21st Century Information Security. *ICCWS 2020 15th International Conference on Cyber Warfare and Security*, 57-66. DOI: 10.34190/ICCWS.20.085
- Buckingham, D. (2019). Teaching media in a ‘post-truth’ age: fake news, media bias and the challenge for media / digital literacy education / La enseñanza mediática en la era de la posverdad: fake news, sesgo mediático y el reto para la educación en materia de alfabetización mediática y digital. *Culture and Education/Cultura y Educación*, 31(2), 213–231. <https://doi.org/10.1080/11356405.2019.1603814>
- Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Anyone can become a troll: Causes of trolling behavior in online discussions. In Lee, C. P., & S. Poltrok (Eds.), *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 1217-1230). Association for Computer Machinery. <https://doi.org/10.1145/2998181.2998213>
- Chesney, R., & Citron, D. (2019). Deepfakes and the New Disinformation War: The Coming Age of Post-Truth Geopolitics. *Foreign Affairs*, 98, 147.
- Drolsbach, C. P., & Pröllochs, N. (2023). Diffusion of community fact-checked misinformation on Twitter. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), 1–22. <https://doi.org/10.1145/3610058>
- Fagni, T., Falchi, F., Gambini, M., Martella, A., & Tesconi, M. (2021). TweepFake: About detecting deepfake tweets. *PLoS ONE*, 16(5), e0251415. DOI: 10.1371/journal.pone.0251415
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communication of the ACM*, 59(7), 96-104. <https://doi.org/10.1145/2818717>

- Gelfert, A. (2018). Fake news: A definition. *Informal Logic*, 38(1), 84-117.
<https://doi.org/10.22329/il.v38i1.5068>
- Global Witness. (2024, December 17). What happened on TikTok around the annulled Romanian presidential election? An investigation and poll. *Author*. Retrieved from <https://globalwitness.org/en/campaigns/digital-threats/what-happened-on-tiktok-around-the-annulled-romanian-presidential-election-an-investigation-and-poll/>
- Gunther, R., Beck, P. A., & Nisbet, E. C. (2018). *Fake News May Have Contributed to Trump's 2016 Victory*. Ohio State University.
<https://assets.documentcloud.org/documents/4429952/Fake-News-May-Have-Contributed-to-Trump-s-2016.pdf>
- INSCOP. (2025). January 2025 – INSCOP Research opinion poll, commissioned by Funky Citizens. Romania between national and European in the era of disinformation. Economic patriotism, values and conspiracies. *Author*. Retrieved from <https://www.inscop.ro/en/january-2025-inscop-research-opinion-poll-commissioned-by-funky-citizens/>
- JURIDICE.ro. (2024, April 2). Ghid privind identificarea materialelor de tip deepfake. Retrieved from <https://www.juridice.ro/731663/ghid-privind-identificarea-materialelor-de-tip-deepfake.html>
- Lev-On, A. (2017). The third-person effect on Facebook: The significance of perceived proficiency. *Telematics and Informatics*, 34(4), 252-260.
<https://doi.org/10.1016/j.tele.2016.07.002>
- Mankoo, S. S. (2023). DeepFakes-The Digital Threat in the Real World. *Gyan Management Journal*, 17(1), 71-77. <https://doi.org/10.48165/gmj.2022.17.1.8>
- Mirsky, Y., & Lee, W. (2021). The Creation and Detection of Deepfakes: A Survey. *ACM Computing Surveys*, 54(1), 1-41. <https://doi.org/10.1145/3425780>
- Narayan, K., Agarwal, H., Mittal, S., Thakral, K., Kundu, S., Vatsa, M., & Singh, R. (2022). DeSI: Deepfake Source Identifier for Social Media. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 2858-2867).
- Newman, N. (2025). Overview and key findings of the Digital News Report. *Reuters Institute*. Retrieved from <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2025/dnr-executive-summary>
- News.ro. (2024, July 10). Sondaj INSCOP la comanda News.ro – Mai mult de 70% dintre români consideră că sunt expuși știrilor false/ Peste 48% cred că în lume se întâmplă multe lucruri foarte importante, despre care oamenii obișnuiți nu sunt niciodată informați. *Author*. Retrieved from <https://www.news.ro/cultura-media/sondaj-inscop-comanda-news-mult-70-dintre-romani-considera-expusi-stirilor-false-pest-48-cred-lume-intampla-multe-lucruri-foarte-importante-oamenii-obisnuiti-niciodata-informati-1922400010002024070821661752>

- O'Carroll, L. (2024, March 26). Hire factcheckers to fight election fake news, EU tells tech firms. *The Guardian*. Retrieved from <https://www.theguardian.com/media/2024/mar/26/tech-firms-poised-to-mass-hire-factcheckers-before-eu-elections?>
- Radu, R.-N. (2024, June 17). Digital News Report. Romania. *Reuters Institute*. Retrieved from <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2024/romania>
- Rini, R. (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal*, 27(2), E-43-E-64. <https://doi.org/10.1353/ken.2017.0025>
- Rochlin, N. (2017). Fake news: belief in post-truth. *Library Hi Tech*, 35(3), 386-392. <https://doi.org/10.1108/lht-03-2017-0062>
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *ACM SIGKDD Explorations Newsletter*, 19(1), 22-36. <https://doi.org/10.1145/3137597.3137600>
- Soll, J. (2016, December 18). The Long and Brutal History of Fake News. *Politico Magazine*. Retrieved from <https://www.politico.com/magazine/story/2016/12/fake-news-history-long-violent-214535/>
- Stănescu, G. C. (2024). Fake News, Bots, and Influencers: The Impact of Social Media on Romania's 2024 Elections. *Social Sciences and Education Research Review*, 11(2), 361-366. <https://doi.org/10.5281/zenodo.15258337>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. *Digital Journalism*, 6(2), 137-153. <https://doi.org/10.1080/21670811.2017.1360143>
- Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2020). Deepfakes and beyond: A Survey of face manipulation and fake detection. *Information Fusion*, 64, 131-148. <https://doi.org/10.1016/j.inffus.2020.06.014>
- Winston, B., & Winston, M. (2020). *The Roots of Fake News. Objecting to Objective Journalism* (1st ed.). Routledge, London. <https://doi.org/10.4324/9780429032264>