CZU: 004.8:659.3:316.776.23 **https://doi.org/10.59295/sum11(7)2026_23**

ECHO CHAMBERS AND FILTER BUBBLES IN THE AGE OF ARTIFICIAL INTELLIGENCE: EFFECTS ON PUBLIC DISCOURSE AND THE CHALLENGES FACING DELIBERATIVE DEMOCRACY

Violeta STRATAN ILBASMIŞ,
Moldova State University

This article examines the impact of Artificial Intelligence (AI) on digital public discourse, focusing on the amplification of echo chambers and filter bubbles. While social networks initially promised inclusive debate and broader access to information, they have increasingly contributed to polarization and selective exposure. AI-driven algorithms personalize content, reinforcing existing beliefs and shaping micro-publics, thereby transforming AI into an epistemic actor. The study analyses the conceptual distinctions between echo chambers and filter bubbles and explores their interaction with algorithmic systems. It highlights the risks posed to democratic deliberation, including fragmentation, radicalization, and the spread of disinformation. The findings emphasize the need for digital literacy, algorithmic accountability, and effective regulation to preserve informational pluralism and support a functional deliberative democracy.

Keywords: *Artificial Intelligence (AI), digital public discourse, echo chambers, filter bubbles, polarization, micro-publics, democratic deliberation, digital literacy.*

CAMERELE DE ECOU ȘI BULELE DE FILTRARE ÎN ERA INTELIGENȚEI ARTICIALE: EFECTE ASUPRA DISCURSULUI PUBLIC ȘI PROVOCĂRILE DEMOCRAȚIEI DELIBERATIVE

Acest articol examinează impactul Inteligenței Artificiale (IA) asupra discursului public digital, cu accent pe amplificarea camerelor de ecou și a bulelor de filtrare. Deși rețelele sociale au fost inițial percepute ca instrumente pentru dezbatere incluzivă și acces extins la informație, ele au contribuit tot mai mult la polarizare și expunere selectivă. Algoritmii bazați pe IA personalizează conținutul, consolidând convingerile existente și formând micro-public, transformând IA într-un actor epistemic. Studiul analizează delimitările conceptuale dintre camerele de ecou și bulele de filtrare și interacțiunea acestora cu sistemele algoritmice. Sunt evidențiate riscurile asupra deliberării democratice, inclusiv fragmentarea, radicalizarea și dezinformarea, subliniind necesitatea alfabetizării digitale, responsabilității algoritmice și reglementării eficiente.

Cuvinte-cheie: *Inteligența Artificială, discurs public digital, camere de ecou, bule de filtrare, polarizare, micro-publicuri, deliberare democratică, alfabetizare digitală.*

Introduction

The emergence and subsequent development of social networks were initially perceived as an opportunity to broaden access to information and to facilitate inclusive debate within the online environment. The evolution of digital media, alongside the expanding influence of social networking platforms, was accompanied by the promise of enhancing digital public discourse and contributing to the consolidation of a genuinely deliberative democracy. However, empirical observations suggest that, rather than fostering mutual understanding through constructive dialogue and reasoned argumentation, social networks have more frequently contributed to the polarization of opinions.

The online environment and social networks have transformed the way individuals access information and participate in public discourse. Things took a radical and rapid turn with the advent of AI. Not without reason, in recent years more and more researchers in media and socio-political studies have turned their attention to examining the effects of Artificial Intelligence on the formation of public discourse in the online sphere. The scale at which AI influences everyday life—and especially the digital public space - its unpredictability, as well as the risks it entails, prompt us to ask what possible mechanisms of regulation and control of AI in the digital sphere might look like.

Beyond its intrinsically unknown nature, opinions regarding the effects of AI are diverse. Some researchers appear highly optimistic about its role in transforming digital public discourse, while others remain skeptical about the positive effects AI might bring. One thing, however, is certain: through recommendation algorithms, personalization systems, and automated content generation, artificial intelligence has contributed to the accelerated digitalization of public discourse. Yet digitalization does not always - or in all circumstances - bring benefits. In this context, echo chambers and filter bubbles have returned to the attention of researchers, as they have become central phenomena for understanding how opinions are formed and circulated in the digital public sphere, amplified by the effects of AI.

The main problem posed by social networks in the age of AI is selective access to information and the polarization of public opinion. Through their algorithmic policies, phenomena such as manipulation, disinformation, and propaganda have acquired new dimensions by delivering personalized content, thereby amplifying or reinforcing the existence of echo chambers. A well-known example in this regard is the Cambridge Analytica scandal, which demonstrated the influence and negative impact that social networks, through their algorithmic policies, can have on the deliberative process in the United States.

Social networks have amplified the voices of disadvantaged groups, yet these are often exploited by opportunistic political actors to disseminate manipulative content. Dominant platforms claim to support digital and deliberative democracy, but through online plebiscites and campaigns, they often contribute to the promotion of populism and authoritarianism.

Echo chambers and filter bubbles: conceptual delimitations

Echo chambers describe informational environments in which individuals are predominantly exposed to opinions that confirm their existing beliefs - environments where people encounter only views similar to their own and avoid opposing perspectives. Echo chambers emerge when their members begin to discredit outsiders and place trust exclusively in those within their own circle of contact. Those outside the chamber are perceived as lacking credibility and, at times, even as hostile.

Echo chambers act as a catalyst for digital behavior, thereby also becoming a catalyst for the online process of disinformation. Users tend to engage in discussions with like-minded individuals, which leads to the formation of homogeneous opinion groups [1, 2].

The theory of online polarization and belief confirmation has been extensively discussed by researcher Cass Robert Sunstein, who argues that in such environments, opinions are repeated and mutually reinforced, leading to their consolidation and radicalization [3].

The phenomenon of echo chambers has also been interpreted from a socio-psychological perspective. In the context of Social Identity Theory, individuals prefer to feel a sense of belonging to a social group and to experience the value and emotional significance attached to that membership. At the same time, people tend to evaluate their group through comparison with other groups, seeking a positive social identity. According to this theory, individuals favor their own group, and prejudice, discrimination, and intergroup conflicts are not driven solely by economic or realistic factors (such as competition over resources), but also by the psychological need for a positive social identity [4, p. 40].

An increasing number of studies describe echo chambers as a threat to intellectual debate and the freedom to engage with differing opinions. Nevertheless, the results of several empirical studies focusing on social networks on X (formerly Twitter) have found that the majority of users are not extremely polarized and tend to share moderate views. However, 1% of the most active users post and disseminate highly polarized and extreme content [5].

Data from another study conducted on X (based on a sample of 40,000 users divided according to political ideology) suggest that moderate opinions nevertheless tend to disappear over time, while those in the middle begin to adopt viewpoints more aligned with their community. The study indicates that opinions are not defined solely by individual preferences but are shaped by the social echo within users' networks [6].

If one seeks to establish a correlation with the field of psychology or to provide a strictly psychological interpretation, it is pertinent to refer to the experiment conducted by Verplanck (1955), which was carried

out prior to the advent of social networks. The study demonstrated that validating and paraphrasing an interlocutor's opinions leads to an increased frequency in the expression of those opinions, highlighting a mechanism of message amplification that anticipates the functioning of echo chambers [7].

A study conducted in Romania set out to demonstrate how algorithms and political actors use echo chambers for manipulation, political capital, and the spread of disinformation. The research analyzed a Facebook group with approximately 93,800 members, centered around a political figure. The findings showed that several recurring mechanisms operate within the echo chamber: the creation of a strong sense of community ("us vs. them"); repetitive messages that reinforce group beliefs; the consolidation of a shared identity (nationalism, patriotism); the rejection of opposing viewpoints; emotional and manipulative language, including alarmist terms; and the rapid shifting of topics, which reduces the capacity for critical analysis [8].

In conclusion, it can be argued that echo chambers shape cognitive processes by reinforcing group affiliation and social identification; through membership within a collective, individuals tend to align their beliefs and judgments with the prevailing framework of that group.

Filter bubbles, a concept popularized by Eli Pariser, refer to the algorithmic filtering of information that limits users' access to alternative perspectives [9]. In the age of AI, the two phenomena overlap and reinforce one another. Filter bubbles personalize information based on users' online behavior. Digital platforms collect data about users and then use this data to display content that appears relevant to them. As a result, content becomes highly personalized, ensuring the operational model or core principle by which platforms function-namely, targeted advertising and user engagement.

Researchers of filter bubble theories, such as Cass Robert Sunstein and Eli Pariser, argue that the internet creates filter bubbles in which we do not encounter opposing opinions. In theory, if the problem were only a lack of exposure, simply being exposed to opposing viewpoints would "break the bubble". But reality is different. Even when members of an echo chamber hear opposing arguments - for example, on climate change, political ideologies, or sensitive social issues - they often reject them, citing distrust in institutions or mainstream media.

The greatest danger arises when the public relies on digital sources for information, where the risk of falling prey to disinformation and manipulation is high. This risk is even greater when individuals are already members of these echo chambers, in which they place complete trust, while legitimate external sources hold little interest for them. The amplification of isolation and the redirection of attention solely within echo chambers contribute to the reduction of dialogue and deliberation in a democratic state - a danger that researchers in the field have highlighted.

Within the broader debate on echo chambers, several scholars also invoke the concept of an "epistemic bubble", defined as a condition in which individuals are not exposed to dissenting viewpoints - often as a consequence of algorithmically personalized information feeds that limit access to alternative perspectives [10].

A less common term is "cyber-balkanization", which refers to the fragmentation of the internet into "digital tribes" or small, ideologically homogeneous echo chambers, where users primarily interact with like-minded individuals and avoid opposing opinions. The term cyber-balkanization was used in 1997 by Marshall Van Alstyne and Erik Brynjolfsson and derives from the concept of "balkanization" (the division of a region into smaller, hostile units) to describe how online spaces are split into isolated interest groups [11]. Jan van Dijk describes the filter bubble phenomenon as a "self-selecting informational prison" [12, p. 321].

A study conducted in 2020 by a team of researchers in the United States provided a nuanced demonstration of how social networks influence the diversity and ideological orientation of information consumption, showing that the effects - whether polarization or pluralism - depend on the platform used and the measurement methods, rather than reflecting a uniform phenomenon of echo chambers or filter bubbles. The authors analyzed a panel dataset containing over four years of web browsing history from a representative sample of nearly 200,000 adults in the U.S. Using data at the individual level allowed them to observe changes in the consumption of information sources based on digital platform usage. The study offers a conceptual

clarification regarding echo chambers and filter bubbles: these are not simple "yes/no" categories but rather complex variations in information consumption that depend on the platform, algorithm characteristics, and user preferences [13].

If echo chambers control how people think together with others, filter bubbles control what those people see - or what they are allowed to see. They can operate independently, but the most problematic - and dangerous - scenario occurs when they overlap: algorithms direct individuals toward homogeneous communities, and those communities then enclose them within their own chambers. Echo chambers and filter bubbles are two distinct phenomena, yet they are often confused.

Empirical research indicates that online platforms and social networking services utilize users' behavioral data to curate content that corresponds closely to their demonstrated preferences, thereby limiting exposure to divergent viewpoints and reinforcing preexisting attitudes. Such practices may contribute to informational isolation and diminish individuals' capacity to engage with a plurality of perspectives. In this context, algorithmically personalized news feeds can generate dynamics analogous to echo chambers, filter bubbles, and cyber-balkanization [14].

From this perspective, the "global village" conceptualized by Marshall McLuhan appears, in the digital era and the age of artificial intelligence, to be fragmenting into smaller, ideologically homogeneous enclaves that at times display tendencies toward communicative or political separatism.

The role of Artificial Intelligence in amplifying echo chambers and filter bubbles

This article seeks to address the following research questions: What impact does artificial intelligence have on the formation and persistence of echo chambers and filter bubbles? To what extent does AI amplify or mitigate these phenomena? How do AI-driven algorithms operate within the digital democratic sphere? Do they function as catalysts for a genuinely deliberative democracy, or do they instead constrain such deliberation? Finally, what role do AI algorithms play in shaping the scale and speed with which AI has entered our lives compel us to reflect on its potential effects and the role it plays. Existing research indicates that Large Language Models (LLMs) have taken on characteristics of echo chambers. These systems tend to align with users' expressed opinions rather than offering alternative or critical perspectives. The findings show that all models exhibit a significant tendency to confirm users' opinions, although the intensity of this effect varies from one model to another. This tendency is problematic because it can limit users' exposure to differing viewpoints, reinforce existing beliefs, and promote polarization and cognitive bias. The authors emphasize the need to develop mechanisms that mitigate this "echo chamber" effect in conversational AI systems [15].

Creating an echo chamber effect through AI - or even within AI itself - can lead individuals to see only content that confirms their beliefs and aligns with their own perspectives. Algorithmic feedback anticipates user behavior, makes predictions, and, as a result, can deliver increasingly homogeneous content. This homogenization of information may contribute to the gradual radicalization of discourse through repeated exposure to extreme or emotionally charged positions.

Chatbots and large language models can validate preexisting beliefs and generate hyper-personalized content, contributing to the creation of "micro-echo chambers" at the individual level. In this scenario, the risk increases exponentially because people tend to perceive AI as an epistemic authority; it is no longer merely an intermediary of information but becomes an epistemic actor. Through interactions with LLMs, information is personalized in a way that is characterized by excessive cognitive confirmation.

From a psychological perspective, we can observe the comfort people derive from these "micro-chambers." The psychological dimension of AI algorithms suggests the emergence of a harmonious relationship between emotion, identity, and the information consumed by individuals. The danger escalates when we realize that AI models decide what is relevant, credible, or worthy of visibility. In this context, there is a shift of authority - from journalists, experts, and traditional gatekeepers toward opaque algorithmic systems.

AI enables the personalization of information at the individual level rather than the group level, leading to the emergence of unique "individual bubbles." As a result, the public effectively becomes a collection of micro-publics. Language models, by reformulating users' opinions, can reinforce their beliefs - an instance of AI-assisted confirmation bias.

The diagram below provides a visual representation of this triadic relationship, illustrating the points of intersection among the three phenomena and highlighting how their convergence may generate informational isolation and intensify polarization.

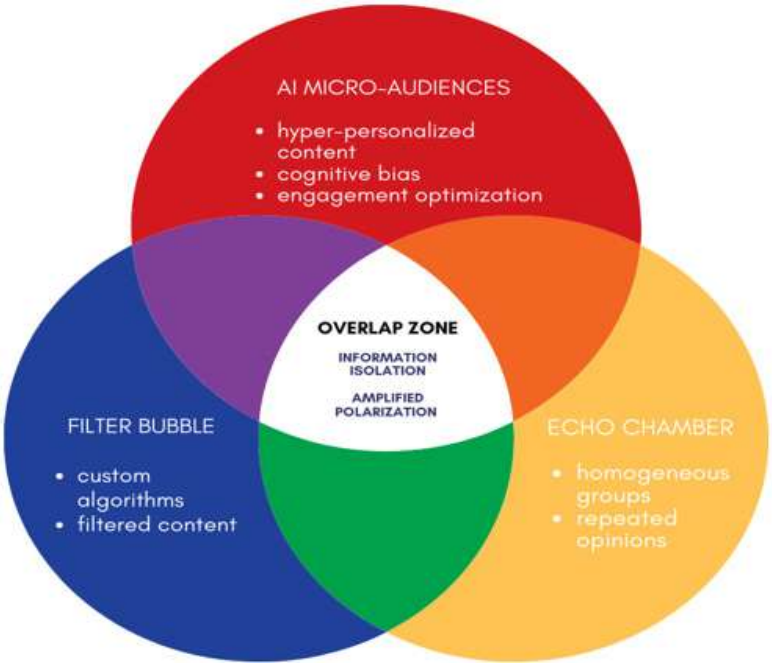


Figure 1. Overlap of filter bubbles and echo chambers in the AI era
Source: Developed by the author

Effects on public discourse

Echo chambers and filter bubbles - amplified by AI effects or even by AI itself - have a significant impact on digital public discourse. The fragmentation of the public sphere into isolated ideological communities, the decline of deliberative dialogue and compromise, the radicalization of opinions, the normalization of extreme discourse, and the diminishing trust in common sources of authority (press, science, institutions) are just some of these effects.

Public discourse in the digital environment becomes driven by emotions rather than critical thinking, more conflictual, and less oriented toward consensus or rational argumentation. All of these factors have major implications for democratic deliberation.

There are also authors who are skeptical about the explanatory power of the terms analyzed in this article and critique the use of the metaphors "echo chambers" and "filter bubbles" in discussions about social networks and polarization. They argue that these concepts are vague metaphors, used both in academic and media contexts without clear definitions. In their view, these metaphors are inadequate for explaining polarization and disinformation, as empirical evidence does not support them as dominant technological phenomena; instead, the real issue is social and political polarization, which goes beyond technology itself [16].

Mostafa M. El Bermawy recounts how, in 2016, prior to the U.S. presidential elections, an analysis of social media followers and engagement surprised him: Donald Trump's numbers were much higher than Hillary Clinton's, yet many people in his liberal circle did not see this in their feeds. The author argues that the internet, once viewed as a "global village," has transformed into segregated digital islands characterized by intense content personalization. This personalization creates increasingly isolated online experiences.

He also criticizes Facebook for refusing to acknowledge its role as a news source and for prioritizing engagement and advertising over informational diversity. El Bermawy argues that recommendation algorithms serve users' comfort rather than exposing them to diverse viewpoints. However, this reduces the public's capacity for constructive dialogue and has negative implications for democracy [17].

Through the passive selection of information, the limitation of opinion diversity, and the reinforcement of existing beliefs, filter bubbles affect public discourse. Similarly, through social interaction among individuals with similar opinions, the reinforcement and radicalization of beliefs, echo chambers impact digital public discourse. These changes have an irreversible effect on democratic deliberation because they reduce the common space for argumentation, fragment the public, and contribute to diminishing mutual trust.

Another feature of recommendation algorithms must be noted: they prioritize engagement over opinion pluralism. Because conflictual content is more "profitable," rational deliberation is algorithmically disadvantaged.

Moreover, generative models can rapidly produce large volumes of content aligned with a particular narrative, increasing the density of similar messages in an already filtered ecosystem. According to internal disclosures from Facebook, algorithms are designed to amplify polarizing content and disinformation because it generates more traffic and, consequently, advertising revenue. Users, including relatively moderate ones, are "pushed" toward increasingly polarizing content to drive visibility and engagement [18]. Algorithmic influence can distort public opinion and affect political outcomes and the democratic process by promoting disinformation and division.

High-profile cases such as Brexit, climate change, or immigration policy show how individuals find highly polarized online groups in which discussions take place. However, the problem does not necessarily lie in what information is accessible, but in who is considered trustworthy within these groups [19].

Other studies emphasize the idea of "framing as a bridge between paradigms, as we move from an era of mass communication to one of echo chambers, personalized information, and micro-targeting in the new media environment" [20]. In digital democracies, these phenomena can affect civic information processes and collective decision-making. Citizens no longer operate with a shared set of facts but with algorithmically shaped parallel realities.

Conclusions

In the era of artificial intelligence, echo chambers and filter bubbles should not be understood merely as unintended by-products of technological innovation, but rather as outcomes shaped by specific design choices and underlying economic imperatives. The extent to which AI influences public discourse depends largely on the equilibrium established between algorithmic optimization and social responsibility. Users who possess higher levels of digital literacy are generally better equipped to navigate the online environment and to encounter heterogeneous viewpoints. By contrast, individuals lacking such competencies may develop a form of "tunnel vision", partly as a consequence of the cognitive overload and stress associated with navigating complex information ecosystems.

The empirical evidence discussed above indicates that, although echo chambers are indeed present within digital communication spaces, not all online interactions occur within such insulated environments. Nonetheless, these structures appear to exert a disproportionate influence on the circulation of polarized viewpoints. By narrowing the spectrum of perspectives to which individuals are exposed, they may shape the quality of public debate and, consequently, warrant sustained scholarly and regulatory attention - even in contexts where platforms formally endorse freedom of expression.

In this context, legislative intervention appears necessary to reform existing frameworks governing data ownership and to enhance competitiveness within the digital marketplace. Reconfiguring social networks into spaces conducive to genuine deliberation - rather than mere confrontation - constitutes a central challenge for contemporary democratic states. Online echo chambers pose a significant risk to democratic societies, and addressing their effects requires a multifaceted approach, including strengthened media education, the development of digital literacy and critical thinking skills, greater exposure to diverse viewpoints, and more robust regulatory oversight to ensure platform accountability. Furthermore, improving journalistic standards and enhancing technological transparency are essential, particularly with regard to clarifying how personalized news content is generated, curated, and disseminated.

In the era of artificial intelligence, the equilibrium among freedom of expression, economic profit, and the public interest has become increasingly fragile. The preservation of this balance depends on the effective implementation of the measures outlined above, in a manner that safeguards democratic values while accommodating technological innovation.

Bibliography:

1. BURNS, A., *It's Not the Technology, Stupid: How the 'Echo Chamber' and 'Filter Bubble' Metaphors Have Failed Us*, *Disponibil*: https://snurb.info/files/2019/It%e2%80%99s%20Not%20the%20Technology,%20Stupid.pdf [Accesat: 31.01.2026]
2. CACCIATORE, M. A., SCHEUFELE, D. A., IYENGAR, S. *The end of framing as we know it ... and the future of media effects.* Mass Communication and Society, 19(1). 7-23, 2016
3. EL-BERMAWY, M.M., *"Your Echo Chamber is Destroying Democracy" Disponibil:* https://www.wired.com/2016/11/filter-bubble-destroying-democracy/ [Accesat: 12.02.2026]
4. FLETCHER, R. *Selective Exposure in the Digital Age: Fact vs Perception*. *Journal of Communication*, 68(3), 512–534, 2018.
5. FORTUNATO, P. '*Social media', market power and the health of democracy*, *Disponibil*: https://www.socialeurope.eu/social-media-market-power-and-the-health-of-democracy [Accesat: 31.01.2026]
6. JAMIESON, K. H., CAPPELLA, J. N. (2008). *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment.* Oxford University Press, 2008, p.320
7. KITCHENS, B., JOHNSON, S. L., GRAY, P., **„***Understanding Echo Chambers and Filter Bubbles: The Impact of Social Media on Diversification and Partisan Shifts in News Consumption" MIS Quarterly* Disponibil: https://www.darden.virginia.edu/sites/default/files/inline-files/05_16371_RA_KitchensJohnsonGray%20Final_0.pdf [Accesat: 13.01.2026]
8. MOSLEY, T., *How Social Media Echo Chambers Drown Out the Voices in the Middle, Disponibil:* https://www.kqed.org/news/11703717/how-social-media-echo-chambers-drown-out-the-voices-in-the-middle [Accesat: 10.01.2026]
9. NARIN, B., *The examination of personalized news feed in the context of echo chamber effect, filter bubble and cyberbalkanization*, Selçuk İletişim, 2, 11 (2): 232-251 Disponibil: https://dergipark.org.tr/tr/download/article-file/504829 [Accesat: 10.01.2026]
10. NEHRING, J., GABRYSZAK, A., JÜRGENS, P., BURCHARDT, A., SCHAFFER, S., SPIELKAMP, M., STARK, B., *Large Language Models are Echo Chambers* In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), p. 10117–10123, Torino, Italia. ELRA, ICCL. Disponibil: https://aclanthology.org/2024.lrec-main.884.pdf [Accesat: 9.01.2026]
11. NGUYEN, C. Thi, *The problem of living inside echo chambers***.** Disponibil: https://theconversation.com/the-problem-of-living-inside-echo-chambers-110486 [Accesat: 8.01.2026]
12. PARISER, E. (2011). *The Filter Bubble: What the Internet Is Hiding from You.* Penguin Press, 2011
13. SHORE, J., BAEK, J., DELLAROCAS, C. *Twitter is not the echo chamber we think it is.* FRONTIERS. MIT Press. Disponibil: https://sloanreview.mit.edu/article/twitter-is-not-the-echo-chamber-we-…[Accesat: 8.01.2026]
14. STOICA, Ş. E., *Disinformation Dynamics: Unveiling the Impact of Echo Chambers in Shaping Online Public Opinion*, *Bulletin of "Carol I" National Defence University*, Vol. 13-1, 2024
15. SUNSTEIN, C. R. *Republic.com,* 2001*,* p.224
16. SUNSTEIN, C. R. *Risk and reason: Safety, law, and the environment.* Cambridge University Press, 2002, p. 342
17. TAJFEL, H., TURNER, J.C., *An integrative theory of intergroup conflict*. În: Austin, W.G., Worchel, S. (Eds.), The Social Psychology of Intergroup Relations. Brooks/ Cole, Monterey, CA, 1979, p.40
18. VAN ALSTYNE, M., BRYNJOLFSSON, E., *Electronic Communities: Global Village or Cyberbalkans?* Disponibil: https://web.mit.edu/marshall/www/papers/CyberBalkans.pdf [Accesat: 8.01.2026]
19. VAN DIJK, J. *The Network Society,* 2016, p. 321
20. VERPLANCK, W.S. *The control of the content of conversation: reinforcement of statements of opinion.* The Journal of Abnormal and Social Psychology, 51(3), 1955, p. 668.

Data about author:
Violeta STRATAN ILBASMIȘ, PhD, University Lecturer, Faculty of Journalism and Communication Sciences, Moldova State University.
ORCID: 0000-0002-0456-7413
E-mail: violetastratan@gmail.com

Presented: 28.02.2026
Reviewd: 10.03.2026
Accepted for publication: 20.05.2026