ISSN 2078-7324. Вісник Львівського університету. Серія Журналістика. 2026. Випуск 58. С. 263–277
Visnyk of the Lviv University. Series Journalism. 2026. Issue 58. P. 263–277

МЕДІАТЕХНОЛОГІЇ

Visn. Lviv. Univ., Ser. Zhurn. 2026: 58; 263–277 • DOI: http://dx.doi.org/10.30970/vjo.2026.58.13928

УДК: 070:004.8:316.774

FROM CENSORSHIP TO BLIND SPOTS: THE EVOLUTION OF INFORMATION CONTROL IN THE AGE OF ALGORITHMS

Vladyslav Bilozerov
Ivan Franko National University of Lviv
General Chuprynka Str., 49, Lviv, Ukraine, 79044
e-mail: Vladyslav.Bilozerov@lnu.edu.ua
https://orcid.org/0009-0004-4556-8916

Yana Shekeriak-Kushka
Ivan Franko National University of Lviv
General Chuprynka Str., 49, Lviv, Ukraine, 79044
e-mail: Yana.Shekeriak@lnu.edu.ua
https://orcid.org/0000-0002-8187-4062

This article examines the hidden (missing) variable as an informational blind spot in contemporary media systems. It analyzes the shift from direct censorship to algorithmic and AI-driven control of visibility, conceptualized as the AI Curtain. In this regime, information is shaped through ranking, filtering, and predictive suppression rather than explicit bans. The study highlights risks for journalism and public discourse and proposes mitigation mechanisms, including algorithmic transparency, pluralistic access to information, and strengthened human editorial oversight.

Keywords: new media, algorithmic control, artificial intelligence, big data, algorithmic censorship, informational blind spots, AI Curtain, data journalism, media, media propaganda, journalism, mass communication.

Relevance and Problem Statement. Throughout history, information has been recognized as a valuable resource. The informational ecosystem manifested itself in every state, and after a considerable period of time, a significant proportion of the population came to recognize the importance of effective information dissemination. The ability to regulate the dissemination of information can yield numerous opportunities.

Even more, information has become a central resource shaping political decision-making, public discourse, and individual perceptions of reality. The rapid expansion

of digital technologies, open data initiatives, and AI-driven information systems has significantly increased both the volume and accessibility of information. However, this apparent abundance of data has not resulted in greater transparency. On the contrary, it has contributed to the emergence of new forms of informational opacity, where critical elements are systematically obscured, deprioritized, or rendered invisible.

The evolution of information control demonstrates a clear transition from direct, state-imposed censorship toward algorithmic and AI-mediated governance of visibility. In contrast to traditional censorship, which relied on explicit bans and legal restrictions, contemporary information control operates through ranking systems, recommendation algorithms, automated moderation, and predictive suppression. These mechanisms do not prohibit content directly but shape what is seen, amplified, or ignored, creating structural "blind spots" in the information space. Such blind spots function independently of journalistic intent or content quality, and increasingly determine the boundaries of public knowledge.

The growing reliance on open data and large-scale datasets further complicates this dynamic. While open data is often framed as a tool for transparency and democratic accountability, its interpretation and circulation are mediated by closed algorithms and AI systems. As a result, visibility is unevenly distributed, and certain narratives, social groups, or types of information remain underrepresented or excluded altogether. Artificial intelligence amplifies this process by learning from historically filtered data, reinforcing prior exclusions and transforming temporary visibility decisions into long-term structural absences.

For contemporary journalism, and particularly for data journalism, these transformations pose fundamental challenges. Data-driven reporting is commonly associated with objectivity, neutrality, and evidentiary strength. Yet, when the data itself is shaped by algorithmic selection and AI-driven omission, journalists risk reproducing invisible biases and informational gaps. This is especially critical in contexts of political conflict, propaganda, and war, where algorithmic systems may suppress documentation of violence, marginalize non-dominant narratives, or amplify coordinated disinformation.

Thus, the relevance of this study lies in the urgent need to critically examine how big data and artificial intelligence contribute to the production and normalization of informational blind spots. Understanding these mechanisms is essential for assessing the evolving nature of censorship, safeguarding epistemic diversity, and redefining the professional responsibility of journalists and media institutions in AI-mediated information environments.

The purpose of this article is to examine the transformation of information control in contemporary media systems, focusing on the emergence of informational blind spots produced by algorithmic, big data, and AI-driven mechanisms. Particular attention is given to the concept of the AI Curtain as a structural form of soft censorship that governs visibility and access to information without explicit prohibitions.

To achieve this purpose, the article addresses the following objectives:

- to analyze the historical evolution of censorship from direct state-imposed restrictions to algorithmic and AI-mediated forms of visibility control;
- to conceptualize the hidden (missing) variable as an informational blind spot independent of content production or journalistic intent;
- to examine how big data practices and AI systems reinforce, scale, and stabilize these blind spots through ranking, moderation, and informational cannibalization;

- to outline potential approaches to mitigating algorithmic blind spots, emphasizing transparency, pluralistic access to information, and the preservation of human editorial judgment.

Methodology. The study combines general scientific and specialized methods tailored to the analysis of algorithmic influence on information visibility. Descriptive and comparative approaches were used to examine scientific literature, media reports, and journalistic practices, with particular attention to algorithmic ranking, AI moderation, and their role in creating informational blind spots. Comparative analysis enabled evaluation of Ukrainian and international cases, highlighting differences in algorithmic transparency, content visibility, and ethical considerations.

Systems and structural-functional methods were applied to investigate the interplay between technological design, media practices, and public access to information, identifying how algorithmic and AI-driven mechanisms shape what is seen or hidden in digital environments. Generalization and synthesis were employed to formulate conclusions regarding the structural and ethical implications of algorithmic information control, and to propose measures for enhancing transparency, pluralistic access, and human editorial oversight.

Literature review. Early research on information control conceptualized censorship as a direct exercise of state power designed to limit access to undesirable content. Classic media studies emphasize how traditional censorship operated through overt legal restrictions and institutional suppression of speech to shape public perception (e.g., state media control in authoritarian contexts). However, the rise of digital media has prompted scholars to reconsider gatekeeping beyond explicit bans, focusing instead on how visibility is mediated by socio-technical systems.

For example, contemporary research made in The Autonomous University of Barcelona highlights the role of algorithmic gatekeeping, where recommender systems and ranking algorithms on platforms like YouTube and TikTok determine which content is foregrounded or relegated, effectively shaping public discourse in ways that transcend human editorial judgment (Nigar Garajamirli, 2025[1]). These systems prioritize engagement and business incentives, redistributing attention away from diverse or critical perspectives and undermining democratic ideals of equal participation in the public sphere.

With the rise of digital platforms, the nature of censorship has shifted. Algorithmic gatekeeping – where ranking, personalization, and recommendation systems determine what users see – creates conditions in which content is not banned but rendered obscure or invisible due to system-level optimization for engagement and commercial priorities. These dynamics contribute to structural informational blind spots that operate independently of formal censorship and human editorial decisions Studies in communication and media theory also indicate that algorithmic structures create informational asymmetries – not through overt censorship, but by controlling visibility via opaque and proprietary mechanisms (Dan Valeriu Voinea. 2025[2]; Joseph, J., Babu, J., Rajasekar, F. V., Philip, S. R., Thomas, R. та V,

[1] Garajamirli, N. (2025), "Algorithmic Gatekeeping and Democratic Communication: Who Decides What the Public Sees?", *European Journal of Communication and Media Studies*, Vol. 4, No. 3, p. 1–20. URL: https://eu-opensci.org/index.php/media/article/view/554?utm (date of access: 1.12.2025)

[2] Dan, Valeriu Voinea (2025), "Reconceptualizing gatekeeping in the age of artificial intelligence: A theoretical exploration of artificial intelligence-driven news curation and automated journalism", *MDPI*. URL: https://www.mdpi.com/2673-5172/6/2/68?utm (date of access: 02.12.2025)

R. P. 2025[3]). This transformation complicates traditional gatekeeping theory by introducing non-human agents into the processes that determine what audiences see and know.

Artificial intelligence further entrenches these patterns, as models trained on historically filtered and biased data reproduce exclusions, narrow the informational horizon, and shape discourse through predictive suppression and feedback loops. Research on AI-driven information ecosystems shows that algorithmic interventions can amplify misinformation, reinforce dominant narratives, and obscure less engaging but socially significant voices, underscoring algorithmic control as a form of soft censorship in networked media environments.

AI's integration into these systems introduces additional layers of complexity. Research on AI-driven disinformation demonstrates that generative models and engagement-optimization algorithms can amplify false or manipulated content, distorting public discourse and exacerbating informational inequalities. These AI components not only shape what information spreads but also embed biases from training data into the structures that govern visibility, making algorithmic influence a central concern for democratic resilience. (Romanishyn A., Malytska O., Goncharuk V. 2025[4] ; Shaojing Sun, Zhiyuan Liu, David Waxman, 2024[5]).

The literature reflects a shift from understanding censorship as legal suppression toward recognizing algorithmic and AI-mediated control of visibility as a form of soft censorship. However, despite this growing body of work on digital media ecosystems, there remains a significant gap in research that explicitly theorizes and empirically investigates how algorithmic and AI-driven mechanisms produce and sustain informational blind spots – the conditions in which certain perspectives, events, or narratives become systematically obscured even without direct censorship. This gap underscores the relevance of the present study, which seeks to conceptualize the hidden variable as an informational blind spot and analyze how big data and AI complicate the distribution and visibility of information in contemporary media systems

Main Body of the Research. Censorship is a practice utilized by governments across various ideological and historical periods. It is a method that can be considered both simple and widespread. It is reasonable to hypothesize that people would not revolt or demonstrate their concerns if they were simply unaware of the problem. Censorship comes in many forms and can have direct or indirect (also known as soft censorship) approaches. The period known as the 'prime of direct censorship' is generally considered to have begun in the 20th century, coinciding with the rise of authoritarian regimes.

It is evident that both the National Socialist German Workers' Party (NSDAP) and the Soviet Union, upon attaining authority, proceeded to dismantle any resistance to their rule. Newspapers were closed down, and those voices that were not deemed acceptable were silenced. In the aforementioned countries, governments functioned as the preeminent

[3] Joseph, J. et al. (2025), "Digital silence: how algorithmic censorship undermines academic freedom in the Global South", *Frontiers*. URL: https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1640244/full?utm(date of access: 03.12.2025).

[4] Romanishyn, A., Malytska, O., Goncharuk, V. (2025), "AI-driven disinformation: policy recommendations for democratic resilience", *PMC Home*. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12351547/?utm (date of access: 03.12.2025).

[5] Sun, S., Liu, Z., Waxman, D. (2024), "A dynamical measure of algorithmically infused visibility", *arXiv.org*. URL: https://arxiv.org/abs/2412.04735?utm (date of access: 5.12.2025).

purveyors of information, with hundreds of them available to the public (even though there were many). Consequently, people were only exposed to the information that had been deemed acceptable by the relevant authorities, which emphasized the accomplishments of the favorable image of the German nation and the newly established Soviet era. This censorship was not only limited to the media but also extended to cultural spheres.

The Nazi German ideas and propaganda were destroyed on the battlefields of WWII but the Soviets did not suffer that. A salient exemplification of post-WWII censorship in the USSR pertains to the theme of veterans and individuals living with disabilities. The majority of researchers concur that the Soviet Union suffered an estimated 30 million fatalities during the Second World War. Consequently, a significant number of individuals from both military and civilian backgrounds sustained physical disabilities and amputations. However, it is not possible to provide an exact figure as this information was classified and not publicly available. The absence of media coverage and public appearances by veterans with such injuries is indicative of a policy that, in the words of a Soviet joke, encapsulates a half-truth: "There are no invalids in the USSR" (Max Elgot, 2020[6]). It is important to note that other significant examples of censorship in the Soviet Union include the Holodomor and the deportation of Crimean Tatars to Central Asia. The policy of renaming cities and destroying or reclaiming cultural heritage in the USSR can also be viewed as an act of censorship. The successful execution of such a campaign in Crimea, which involved the renaming of even the smallest villages with Crimean Tatarian names to nameless Soviet-Russian ones, such as Shchastye and Vesyoloe, is a prime example of this instrument. The implementation of such a censorship policy would have consequences that extend beyond the destruction of connections with previous owners. Indeed, it would also prevent the articulation of numerous questions regarding the history of the location.

In nations where a single source of information is available, the establishment of effective censorship is a relatively straightforward process, particularly during the 20th century. In the context of WWII, the United States of America exercised restraint in the dissemination of information pertaining to the atrocities attributed to the Soviet Union, given the alliance with the communists (Ted Lipien, 2022[7]). However, the combination of this policy with the rise in popularity of communism, precipitated by the defeat of Nazi Germany, engendered a novel predicament: namely, the emergence of pro-Soviet sentiment among the populace. Consequently, a swift and efficacious approach was imperative to address this prevailing attitude. The solution to this problem was a campaign that was later given the name Red Scare. In order to successfully execute such an extensive campaign, the government was required to exercise comprehensive control over the dissemination of information. The US government identified a method for achieving this objective; however, direct non-war censorship was not a viable option. In the United States of America, the First Amendment to the Constitution provides that the government is prohibited from directly censoring the media (although it does not specify what form this censorship might take):

---

[6] lgot, M. (2020), "Soviet attitudes toward disability and the lasting effect on Nagorno-Karabakh", *Humanitarian Aid Relief Trust (HART UK)*. URL: https://www.hart-uk.org/blog/soviet-attitudes-toward-disability-and-the-lasting-effect-on-nagorno-karabakh/ (date of access: 05.12.2025).

[7] Lipien, T. (2022), "Protecting communists from embarrassment: a history of censorship at the Voice of America", *Tadeusz (Ted) Lipien*. URL: https://www.tedlipien.com/2022/11/22/protecting-communists-from-embarrassment-a-history-of-censorship-at-the-voice-of-america/ (date of access: 05.12.2025).

"Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances".[8] At the time, government institutions were not yet strong enough to monitor the compliance of the media with this amendment. Nevertheless, the exercise of control over the media is not the sole means by which this objective can be achieved. The indirect approach to censorship was successfully applied to the information ecosystem of the USA during the Cold War. During the period of the Red Scare, a number of instruments of indirect censorship were employed, including the bribing of journalists and the media, as well as speculation surrounding government funding of the media. Life threats to journalists and the creation of blacklists of influencers were also used as instruments of indirect censorship, and were found to be effective (Matthew Valdespino, 2013[9]).

Algorithms Replacing Censorship. The practice of censorship has been historically linked to the concept of state power, the imposition of legal restrictions, and the overt suppression of speech. Conversely, the early internet was celebrated for its potential to circumvent centralized control and facilitate free expression. However, as the internet evolved – particularly with the rise of social media platforms in the early 2000s – new forms of censorship emerged, embedded not in law but in code. In the early days of the internet, the prevailing emphasis was on openness, chronological ordering, and user-driven discovery. Forums and the nascent form of blogging primarily presented content in linear or user-curated formats. During periods of moderation, this was often characterized by human presence, localized implementation, and a high degree of transparency.

The advent of algorithmic gatekeeping signified a radical paradigm shift. Search engines such as Google have introduced ranking algorithms (e.g., PageRank) that privilege certain sources over others. Subsequently, social media platforms adopted feed-ranking systems, thereby replacing chronological order with relevance-based sorting. These systems effectively transformed platforms into private gatekeepers of public discourse. In contradistinction to conventional censorship, the practice of algorithmic gatekeeping does not entail the direct removal of content. Instead, it exerts a control over the visibility of information, determining which information is amplified and which is relegated to a lesser prominence. This form of censorship is characterized by its subtlety, complexity, and frequent perception as neutral or technical rather than political.

Early social media algorithms were primarily optimized for engagement metrics such as clicks, likes, shares, and time spent on the platform. Content that did not align with these metrics was systematically deprioritized. Consequently, politically sensitive, minority, or non-mainstream perspectives frequently encountered diminished reach, even in instances where they did not violate platform regulations. This phenomenon can be understood as soft censorship, where speech is not banned but rendered effectively invisible. In contrast to state censorship, soft censorship functions without the imposition of formal prohibitions, adherence to due process, or the establishment of accountability mechanisms. It is rare

[8] First Amendment, *Constitution of the United States,* URL: https://constitution.congress.gov/constitution/amendment-1/ (date of access: 05.12.2025).

[9] Valdespino, M. (2013), *American Communism and Cold War Censorship: The Creation of a New American Citizen: thesis for the degree of Bachelor of Arts with Distinction in Political Science,* University of Pennsylvania; advisor J. Green, Philadelphia.

for users to be informed that their content has been down-ranked or filtered, creating an illusion of free expression while structurally limiting its impact.

As the platforms grew in scale, the need for human moderation became apparent. The advent of early algorithmic moderation systems was driven by the necessity to detect a range of online harms, including spam, hate speech, and copyright violations. Subsequent iterations of these systems have also been developed to address the dissemination of misinformation. However, these systems relied on pattern recognition, keyword matching, and probabilistic models that lacked contextual understanding. This situation gave rise to a systematic process of censorship, which had a particular impact on activists engaged in the pursuit of political change, journalists, minority language communities, and content discussing violence, war, or extremism in critical or journalistic contexts. Automated moderation has thus been demonstrated to amplify censorship by prioritizing risk avoidance over the promotion of free expression. Platforms, motivated by legal liability and reputational concerns, demonstrated a preference for false positives (the removal of legitimate speech) over false negatives.

A fundamental attribute of algorithmic censorship is its privatization. Decisions regarding the expression of ideas are made by corporate entities rather than by democratic institutions. The earliest social media companies were responsible for the establishment of community guidelines and algorithmic rules. These were implemented as de facto speech laws, with no public oversight.

Furthermore, these algorithms have historically been, and continue to be, opaque. It is challenging for users to audit, comprehend, or contest the decisions made by these algorithms. This opacity deepens the censorship by removing the possibility of accountability and informed resistance. Censorship becomes entrenched within the very fabric of society rather than being imposed by a visible authority.

Data and Algorithmic Blind Spots. The rise of big data in media has also huge impact on algorithmic blindness. While big data is often promoted as a tool for transparency and democratic accountability, its deployment through opaque algorithms can create structural blind spots that limit public knowledge. Algorithmic systems that rank, filter, or prioritize content often operate as opaque "black boxes," shaping which information reaches audiences and which remains hidden. Errors, biases, or incomplete datasets can amplify informational gaps, creating structural blind spots that function independently of content quality or journalistic intent.

For example, predictive algorithms in finance, such as those analyzed by Cathy O'Neil (2019[10]), use indirect indicators like postal codes to assess creditworthiness, systematically disadvantaging residents of economically marginalized areas. This creates self-reinforcing cycles where excluded groups remain invisible to institutions and opportunities, illustrating how data-driven systems can perpetuate inequality and exclusion. Similar dynamics appear in predictive policing systems like PredPol, where data-driven "high-crime zones" attract increased police presence, producing more arrests and reinforcing the perception of danger in those areas. These cases demonstrate how algorithms can create feedback loops that entrench exclusion or misrepresentation.

[10] О'Ніл, К. (2017), *Зброя математичного знищення: як великі дані збільшують нерівність і загрожують демократії,* Кеті О'Ніл; пер. з англ., Penguin, Лондон, 259 с.

In media contexts, these issues manifest when algorithms prioritize certain content over others. AI-driven recommendation systems, social media feeds, and conversational agents (e.g., chatbots) often operate as "black boxes," with journalists and audiences unable to fully understand how visibility is determined. Incomplete or biased training data can amplify historical exclusions, producing informational blind spots that persist over time. Practices such as anonymization, verification, and transparency in methodology can mitigate these risks, but they remain contingent on human oversight[11].

Thus, while big data enables deeper insight and analysis, it simultaneously introduces new forms of structural censorship. The opaque operation of ranking, filtering, and predictive systems can inadvertently produce "blind spots" in public knowledge, underscoring the need for critical oversight and transparency in AI-mediated media environments.

AI-Driven Biases. As digital platforms grew in scale and complexity, traditional algorithmic systems became inadequate in managing the sheer volume of user-generated content. The implementation of artificial intelligence, particularly machine learning and natural language processing, was proposed as a solution to automate the process of moderation, personalize information flows, and facilitate the detection of harmful content. Nevertheless, this transition did not merely enhance efficiency; it fundamentally transformed the nature of censorship. AI does not merely replicate earlier forms of algorithmic control, but rather serves to intensify them. By facilitating predictive, adaptive, and large-scale intervention in speech, AI-driven systems serve to deepen censorship in less visible ways, even more difficult to contest than algorithmic, and more structurally entrenched.

Early algorithms were characterized by the utilization of relatively fixed rules, ranking formulas, and explicit thresholds. In contrast, AI systems function through probabilistic inference, pattern recognition, and continuous learning. This shift denotes the transition from algorithmic filtering to AI governance of speech.

AI models are capable of not only categorizing content based on its inherent nature, but also on its potential to become harmful, misleading, or controversial. Consequently, content may be preemptively suppressed, before any harm occurs or any rules are formally violated. This predictive logic has the potential to expand the scope of censorship beyond the confines of explicit prohibition. The advent of artificial intelligence has enabled digital platforms to moderate an immense volume of content, including billions of posts, images, and videos, in real time. Although scale is frequently presented as a technical necessity, it comes at the cost of contextual judgment. AI systems have been observed to encounter challenges in processing irony, satire, political nuance, cultural specificity, and the manner in which journalism is framed. The consequence of this phenomenon is the systematic application of censorship, particularly in contexts involving: the practice of reporting on warfare and documenting conflict, expression of opposition to prevailing political beliefs and the active involvement in political activities, minority languages and dialects, and visual evidence of violence used for accountability. AI moderation systems have a tendency to suppress content, as false positives are considered to be safer than allowing controversial content to

[11] Шекеряк-Кушка, Я. В. (2025), "Штучний інтелект у журналістиці даних: можливості, загрози та етичні дилеми", *Вчені записки ТНУ імені В. І. Вернадського, серія: Філологія. Журналістика,* т. 36, № 4, частина 2, УДК 070:004.8+004.6(477) (304-314), DOI: https://doi.org/10.32782/2710-4656/2025.4.2/48.

spread. This phenomenon has a disproportionate impact on journalists, researchers, and civil society activists, whose work frequently involves dealing with sensitive material.

One of the most significant ways in which AI worsens censorship is through the implementation of predictive suppression. AI systems do not merely react to content; they shape user behavior by anticipating and steering future actions. Recommendation algorithms have been developed to deprioritize content that is deemed risky, controversial, or non-conforming to platform norms. This form of governance has been shown to encourage self-censorship. It has been demonstrated that users learn – often subconsciously – to adapt their speech to algorithmic preferences, avoiding topics, language, or imagery that might trigger suppression. This process, when prolonged, has the effect of producing a homogenisation of discourse and a narrowing of the public sphere. This phenomenon is exemplified by Meta's proactive suppression of posts related to the Russo-Ukrainian War after 2022[12].

So we can say that in the context of the Russian-Ukrainian war, algorithmic curation has demonstrated vulnerabilities to manipulation and misinformation. Studies of web search engines during propaganda campaigns reveal that a considerable share of search results can include false or misleading content, with exposure varying by language and location, pointing to systemic biases in algorithmic outputs. Platforms themselves have also been observed to classify posts about Ukraine as "undesirable" or suppress them under misinformation labels, effectively obscuring war-related reporting from wider audiences.

AI-driven moderation systems are particularly vulnerable to political instrumentalization. There is an increasing tendency for states to exert pressure on social media platforms to suppress content that falls under broad categories such as "extremism", "misinformation", or "national security." The utilization of artificial intelligence within platforms facilitates the enforcement of compliance on a large scale, while concurrently ensuring the maintenance of plausible deniability. An exemplification of this was provided by the case of Chinese AI DeepSeek, which was found to be intentionally filtering information related to Taiwan and the anti-communist events[13].

In the context of information warfare, AI censorship mechanisms have the potential to be exploited for the purpose of stifling critical narratives whilst permitting the proliferation of coordinated propaganda[14]. It is evident that sophisticated actors are able to circumvent AI systems, while independent journalists and grassroots voices are disproportionately filtered out. During the Russo-Ukrainian war, the Russians established a disinformation network with news sites utilizing open-source search AI to disseminate false information[15].

---

[12] Kling, J., Poliakoff, S. (2025), "Facebook, the EU and Russia's war: challenges of moderating authoritarian news", *Internet Policy Review.* URL: https://policyreview.info/articles/analysis/russias-war-moderating-authoritarian-news (date of access: 05.12.2025).

[13] Qiu, P., Zhou, S., Ferrara, E. (2025), "Information suppression in large language models: auditing, quantifying, and characterizing censorship in DeepSeek", *arXiv.org.* URL: https://arxiv.org/abs/2506.12349 (date of access: 05.12.2025).

[14] Білозеров, В. В. (2025), "Штучний інтелект як інструмент інформаційної російсько-української війни", *Сучасні медіа: тематично-змістові моделі та трансформації,* Збірник матеріалів звітної наукової конференції за 2024 рік (секція «Журналістики»), Львів, Україна, с. 68-70. https://journ.lnu.edu.ua/wp-content/uploads/2025/09/Zbirnyk-tez-zvitnoi-konferentsii-za-2024-rik.pdf

[15] Sadeghi, M., Blachez, I. (2024), "A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda, NewsGuard's Reality Check", *Substack*. URL: https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global

It is evident that AI systems are characterised by the embedding of normative judgments concerning the definition of acceptable speech. The aforementioned judgments are encoded by private companies, which have been trained on biased datasets and are optimized for corporate risk management as opposed to democratic values. Moreover, private companies possess the capacity to regulate the dissemination of information in a manner analogous to Grok, Elon Musk's AI, and the X/Twitter feed[16]. As artificial intelligence becomes the primary arbiter of visibility and legitimacy online, platforms assume quasi-sovereign power over public discourse. This privatization of normative authority signifies a substantial shift in the governance of speech, characterized by a lack of transparency, accountability, and public consent.

An additional risk introduced by artificial intelligence is the phenomenon of informational cannibalism, wherein AI systems are increasingly trained on data generated by earlier algorithmic and AI-mediated environments. As platforms rely on historical datasets shaped by prior moderation decisions, ranking biases, and censorship practices, AI models do not merely reflect the informational ecosystem – they recursively consume and reproduce it. This self-referential learning process serves to amplify earlier errors, exclusions, and silences. Content that was previously down-ranked, removed, or rendered invisible becomes underrepresented in training data, thus reinforcing its marginalization in future AI decisions. Conversely, dominant narratives are frequently optimized for engagement, safety or political convenience, and are disproportionately preserved and amplified. Consequently, AI systems are imbued with and perpetuate the censorship logics of earlier platforms. Such internal processes of 'cannibalisation'[17] have the effect of limiting 'epistemic diversity' and creating a 'path-dependent information environment' in which alternative perspectives have great difficulty in re-emerging once they have been suppressed. In lieu of rectifying antecedent missteps in the domain of algorithmic governance, there is a risk that artificial intelligence (AI) will serve to entrench them within the technical infrastructure. This recursive dynamic has the effect of transforming temporary moderation choices into long-term structural constraints on knowledge production and public discourse.

The increasing utilisation of AI chatbots as alternatives to conventional search engines has the potential to exacerbate the issues of informational cannibalism and algorithmic censorship[18]. In contradistinction to search engines, which present multiple sources and allow users to compare perspectives, conversational AI systems provide synthesized, singular responses. This shift from pluralistic retrieval to authoritative summarization centralizes interpretive power within the AI system itself. Chatbots are heavily dependent on content that has previously been indexed, ranked, and moderated, a significant proportion of which has already been shaped by earlier algorithmic biases and censorship regimes. Consequently, viewpoints that have been repressed or marginalized – especially those that have previously been subject to downranking or removal – are less likely to manifest in the responses generated by chatbots. The system effectively recycles the outcomes of

[16] Ncube, M. (2025), "Grok AI: A conduit for misinformation in the digital age", *IDRC – Resisting Information Disorder in the Global South*. URL: https://idrc.sun.ac.za/grok-ai-a-conduit-for-misinformation-in-the-digital-age/ (date of access: 5.12.2025).

[17] Belenguer, L. (2022), "AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry - PMC", *PMC Home,* https://doi.org/10.1007/s43681-022-00138-8

[18] Crestodina, A. (2025), "Are AI chatbots replacing search engines? AI vs Google [new research]", *Orbit Media Studio*. URL: https://www.orbitmedia.com/blog/ai-vs-google/ (date of access: 04.12.2025).

earlier visibility decisions, further narrowing the informational horizon. However, given the assumption that a considerable amount of information has already been inaccurately designated as false, this problem is only exacerbated.

Furthermore, conversational interfaces obscure the processes of selection and omission. It is rare for users to be presented with information regarding excluded data, deprioritized sources, or areas of uncertainty. This opacity fosters an illusion of completeness and neutrality while reinforcing dominant narratives. In contexts characterized by political conflict, propaganda, or contested knowledge, the utilization of chatbot-based search engines poses a significant risk of becoming a potent mechanism of epistemic closure. This is due to the propagation of historical biases and prior censorship on a large scale, masked by the facade of helpful automation.

A further dimension of concern in conversational search lies in the incentive structures governing AI chatbots, particularly their optimization for user satisfaction, retention, and perceived helpfulness. In contrast to conventional search engines, which are designed to rank and retrieve external sources, conversational AI systems are engineered to produce coherent, confident, and agreeable responses. This design priority creates pressure to minimize uncertainty, controversy, and cognitive friction – often at the expense of epistemic plurality. In their endeavor to satisfy users, chatbots may be inclined to prioritize simplified, mainstream, or socially acceptable narratives while avoiding contentious or niche perspectives that could provoke discomfort or dissatisfaction. Such alignment incentives subtly shape knowledge access by privileging answers that feel authoritative and reassuring, rather than those that reflect genuine complexity or contestation. It is important to note that, over time, this dynamic risk may reinforce dominant viewpoints and suppress critical or dissenting information, thus further centralizing interpretive authority within AI-mediated systems.

Discussion and Conclusions The challenges posed by algorithmic and AI-mediated censorship are structural rather than incidental, requiring interventions at technical, institutional, and normative levels. The presentation of a solitary solution is inadequate in this case. Rather, a multi-layered response is necessary.

Conversational AI systems should complement pluralistic search rather than replace it. Design choices must prioritize source diversity, explicit citation, and the presentation of competing perspectives. It is imperative that users are exposed not only to synthesized answers but also to the range of interpretations and uncertainties that are characteristic of contested knowledge. Nevertheless, it must be noted that this is a delicate situation, due to the possibility of displaying content that may be potentially damaging, with the intention of offering an alternative perspective.

In order to confront the phenomenon of informational cannibalism, it is imperative that AI systems are trained on datasets that deliberately reintroduce content that has been disregarded, archived, or historically suppressed. It is imperative that periodic audits are implemented in order to identify feedback loops that reinforce earlier censorship decisions. Furthermore, it is essential to correct for path-dependent bias.

It is incumbent upon platforms to provide meaningful explanations for AI-driven visibility decisions, including down-ranking and shadow suppression. It is vital to establish transparent appeal mechanisms and independent oversight to reinstate procedural fairness and accountability in speech governance.

It should be noted that AI systems should not be optimized exclusively for the purpose of eliciting user satisfaction, agreement, or retention. Instead, epistemic responsibility –

that is, accuracy, uncertainty disclosure, and representational balance – must be regarded as core performance metrics, even when this is likely to cause discomfort or complexity.

In domains such as war reporting, political analysis, and historical interpretation, the role of AI should be to assist rather than arbitrate. It is imperative that human editorial judgment remains central to the process, with the utilization of AI tools operating under the parameters of clearly defined and limited mandates.

It was evident that the most challenging aspect of the proposed solutions was not even of a technical nature. The focus of this discussion is on major AI start-ups that voluntarily sacrifice a proportion of their profits with the aim of enhancing the quality of life for individuals around the globe. This constitutes the genesis of efforts to combat the concealment of information in the digital era.

СПИСОК ЛІТЕРАТУРИ

1. Belenguer, L. (2022), "AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry – PMC", *PMC Home,* https://doi.org/10.1007/s43681-022-00138-8
2. Crestodina, A. (2025), "Are AI chatbots replacing search engines? AI vs Google [new research]", *Orbit Media Studio*. URL: https://www.orbitmedia.com/blog/ai-vs-google/ (date of access: 04.12.2025).
3. Dan, Valeriu Voinea (2025), "Reconceptualizing gatekeeping in the age of artificial intelligence: A theoretical exploration of artificial intelligence-driven news curation and automated journalism", *MDPI*. URL: https://www.mdpi.com/2673-5172/6/2/68?utm (date of access: 02.12.2025)
4. First Amendment, *Constitution of the United States,* URL: https://constitution.congress.gov/constitution/amendment-1/ (date of access: 05.12.2025).
5. Joseph, J. et al. (2025), "Digital silence: how algorithmic censorship undermines academic freedom in the Global South", *Frontiers*. URL: https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1640244/full?utm(date of access: 03.12.2025).
6. Garajamirli, N. (2025), "Algorithmic Gatekeeping and Democratic Communication: Who Decides What the Public Sees?", *European Journal of Communication and Media Studies*, Vol. 4, No. 3, p. 1–20. URL: https://eu-opensci.org/index.php/media/article/view/554?utm (date of access: 1.12.2025)
7. Kling, J., Poliakoff, S. (2025), "Facebook, the EU and Russia's war: challenges of moderating authoritarian news", *Internet Policy Review.* URL: https://policyreview.info/articles/analysis/russias-war-moderating-authoritarian-news (date of access: 05.12.2025).
8. Lipien, T. (2022), "Protecting communists from embarrassment: a history of censorship at the Voice of America", *Tadeusz (Ted) Lipien*. URL: https://www.tedlipien.com/2022/11/22/protecting-communists-from-embarrassment-a-history-of-censorship-at-the-voice-of-america/ (date of access: 05.12.2025).
9. Ncube, M. (2025), "Grok AI: A conduit for misinformation in the digital age", *IDRC – Resisting Information Disorder in the Global South*. URL: https://idrc.sun.ac.za/grok-ai-a-conduit-for-misinformation-in-the-digital-age/ (date of access: 5.12.2025).
10. Qiu, P., Zhou, S., Ferrara, E. (2025), "Information suppression in large language models: auditing, quantifying, and characterizing censorship in DeepSeek", *arXiv.org.* URL: https://arxiv.org/abs/2506.12349 (date of access: 05.12.2025).

11. Romanishyn, A., Malytska, O., Goncharuk, V. (2025), "AI-driven disinformation: policy recommendations for democratic resilience", *PMC Home*. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12351547/?utm (date of access: 03.12.2025).
12. Sadeghi, M., Blachez, I. (2024), "A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda, NewsGuard's Reality Check", *Substack*. URL: https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global
13. Sun, S., Liu, Z., Waxman, D. (2024), "A dynamical measure of algorithmically infused visibility", *arXiv.org*. URL: https://arxiv.org/abs/2412.04735?utm (date of access: 5.12.2025).
14. Valdespino, M. (2013), *American Communism and Cold War Censorship: The Creation of a New American Citizen: thesis for the degree of Bachelor of Arts with Distinction in Political Science,* University of Pennsylvania; advisor J. Green, Philadelphia.
15. lgot, M. (2020), "Soviet attitudes toward disability and the lasting effect on Nagorno-Karabakh", *Humanitarian Aid Relief Trust (HART UK)*. URL: https://www.hart-uk.org/blog/soviet-attitudes-toward-disability-and-the-lasting-effect-on-nagorno-karabakh/ (date of access: 05.12.2025).
16. Білозеров, В. В. (2025), "Штучний інтелект як інструмент інформаційної російсько-української війни", *Сучасні медіа: тематично-змістові моделі та трансформації,* Збірник матеріалів звітної наукової конференції за 2024 рік (секція «Журналістики»), Львів, Україна, с. 68-70. https://journ.lnu.edu.ua/wp-content/uploads/2025/09/Zbirnyk-tez-zvitnoi-konferentsii-za-2024-rik.pdf
17. О'Ніл, К. (2017), *Зброя математичного знищення: як великі дані збільшують нерівність і загрожують демократії,* Кеті О'Ніл; пер. з англ., Penguin, Лондон, 259 с.
18. Шекеряк-Кушка, Я. В. (2025), "Штучний інтелект у журналістиці даних: можливості, загрози та етичні дилеми", *Вчені записки ТНУ імені В. І. Вернадського, серія: Філологія. Журналістика,* т. 36, № 4, частина 2, УДК 070:004.8+004.6(477) (304-314), DOI: https://doi.org/10.32782/2710-4656/2025.4.2/48.

REFERENCES

1. Belenguer, L. (2022), "AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry – PMC", *PMC Home,* https://doi.org/10.1007/s43681-022-00138-8
2. Crestodina, A. (2025), "Are AI chatbots replacing search engines? AI vs Google [new research]", *Orbit Media Studio*. URL: https://www.orbitmedia.com/blog/ai-vs-google/ (date of access: 04.12.2025).
3. Dan, Valeriu Voinea (2025), "Reconceptualizing gatekeeping in the age of artificial intelligence: A theoretical exploration of artificial intelligence-driven news curation and automated journalism", *MDPI*. URL: https://www.mdpi.com/2673-5172/6/2/68?utm (date of access: 02.12.2025)
4. First Amendment, *Constitution of the United States,* URL: https://constitution.congress.gov/constitution/amendment-1/ (date of access: 05.12.2025).
5. Joseph, J. et al. (2025), "Digital silence: how algorithmic censorship undermines academic freedom in the Global South", *Frontiers*. URL: https://www.frontiersin.org/journals/communication/articles/10.3389/fcomm.2025.1640244/full?utm (date of access: 03.12.2025).

6. Garajamirli, N. (2025), "Algorithmic Gatekeeping and Democratic Communication: Who Decides What the Public Sees?", *European Journal of Communication and Media Studies*, Vol. 4, No. 3, p. 1–20. URL: https://eu-opensci.org/index.php/media/article/view/554?utm (date of access: 1.12.2025)
7. Kling, J., Poliakoff, S. (2025), "Facebook, the EU and Russia's war: challenges of moderating authoritarian news", *Internet Policy Review.* URL: https://policyreview.info/articles/analysis/russias-war-moderating-authoritarian-news (date of access: 05.12.2025).
8. Lipien, T. (2022), "Protecting communists from embarrassment: a history of censorship at the Voice of America", *Tadeusz (Ted) Lipien*. URL: https://www.tedlipien.com/2022/11/22/protecting-communists-from-embarrassment-a-history-of-censorship-at-the-voice-of-america/ (date of access: 05.12.2025).
9. Ncube, M. (2025), "Grok AI: A conduit for misinformation in the digital age", *IDRC – Resisting Information Disorder in the Global South*. URL: https://idrc.sun.ac.za/grok-ai-a-conduit-for-misinformation-in-the-digital-age/ (date of access: 5.12.2025).
10. Qiu, P., Zhou, S., Ferrara, E. (2025), "Information suppression in large language models: auditing, quantifying, and characterizing censorship in DeepSeek", *arXiv.org.* URL: https://arxiv.org/abs/2506.12349 (date of access: 05.12.2025).
11. Romanishyn, A., Malytska, O., Goncharuk, V. (2025), "AI-driven disinformation: policy recommendations for democratic resilience", *PMC Home*. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC12351547/?utm (date of access: 03.12.2025).
12. Sadeghi, M., Blachez, I. (2024), "A well-funded Moscow-based global 'news' network has infected Western artificial intelligence tools worldwide with Russian propaganda, NewsGuard's Reality Check", *Substack*. URL: https://www.newsguardrealitycheck.com/p/a-well-funded-moscow-based-global
13. Sun, S., Liu, Z., Waxman, D. (2024), "A dynamical measure of algorithmically infused visibility", *arXiv.org.* URL: https://arxiv.org/abs/2412.04735?utm (date of access: 5.12.2025).
14. Valdespino, M. (2013), *American Communism and Cold War Censorship: The Creation of a New American Citizen: thesis for the degree of Bachelor of Arts with Distinction in Political Science,* University of Pennsylvania; advisor J. Green, Philadelphia.
15. lgot, M. (2020), "Soviet attitudes toward disability and the lasting effect on Nagorno-Karabakh", *Humanitarian Aid Relief Trust (HART UK)*. URL: https://www.hart-uk.org/blog/soviet-attitudes-toward-disability-and-the-lasting-effect-on-nagorno-karabakh/ (date of access: 05.12.2025).
16. Bilozerov, V. V. (2025), "Artificial Intelligence as an Instrument of the Information Russia–Ukraine War", *Contemporary media: thematic and content models and transformations,* Proceedings of the annual academic conference for 2024 (Journalism Section). Lviv, Ukraine, pp. 68–70. URL:https://journ.lnu.edu.ua/wp-content/uploads/2025/09/Zbirnyk-tez-zvitnoi-konferentsii-za-2024-rik.pdf (date of access: 06.12.2025).
17. O'Neil, C. (2017), *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*, Cathy O'Neil, translated from English, Penguin, London, 259 p. (date of access: 06.12.2025).
18. Shekeriak-Kushka, Ya. V. (2025), "Artificial Intelligence in Data Journalism: Opportunities, Risks, and Ethical Dilemmas", *Scientific Notes of V. I. Vernadsky Taurida National University, series: Philology. Journalism,* Vol. 36, No. 4, part 2, (304-314), DOI: https://doi.org/10.32782/2710-4656/2025.4.2/48 (date of access: 06.12.2025).

ВІД ЦЕНЗУРИ ДО СЛІПИХ ЗОН: ЕВОЛЮЦІЯ КОНТРОЛЮ ІНФОРМАЦІЇ В ЕПОХУ АЛГОРИТМІВ

Владислав Білозеров
Львівський національний університет імені Івана Франка,
вул. Генерала Чупринки, 49, 79044, Львів, Україна
e-mail: Vladyslav.Bilozerov@lnu.edu.ua
https://orcid.org/0009-0004-4556-8916

Яна Шекеряк-Кушка
Львівський національний університет імені Івана Франка,
вул. Генерала Чупринки, 49, 79044, Львів, Україна
e-mail: Yana.Shekeriak@lnu.edu.ua
https://orcid.org/0000-0002-8187-4062

Статтю присвячено дослідженню трансформації механізмів контролю інформації в умовах цифрових медіа та поширення алгоритмічних і AI-орієнтованих систем. Проаналізовано перехід від прямої цензури до непрямих форм управління видимістю інформації, що реалізуються через алгоритми ранжування, фільтрації та прогнозування контенту. Особливу увагу приділено феномену інформаційних «сліпих зон», які виникають унаслідок алгоритмічного відбору даних без явної заборони або обмеження доступу.

Розкрито поняття прихованої (відсутньої) змінної, що впливає на формування публічного порядку денного та суспільне сприйняття реальності. Запропоновано концепт AI Curtain для опису режиму алгоритмічного контролю, за якого інформація не блокується напряму, а стає малопомітною або маргіналізованою через технічні рішення платформ. Проаналізовано роль великих даних, автоматизованої модерації та рекомендаційних систем у відтворенні упереджень і стабілізації інформаційної нерівності.

Окрему увагу приділено наслідкам алгоритмічного контролю для журналістики та публічного дискурсу. Показано, що журналісти, зокрема в галузі журналістики даних, стикаються з ризиком відтворення невидимих викривлень, зумовлених обмеженим доступом до даних, алгоритмічно сформованими пріоритетами та зворотними петлями навчання AI-систем. Виявлено, що зростання ролі чатботів і AI-посередників як джерел інформації звужує плюралізм інтерпретацій та посилює концентрацію влади над знанням. Окреслено можливі механізми пом'якшення негативного впливу алгоритмічної цензури, зокрема підвищення прозорості алгоритмів, забезпечення множинності джерел інформації, проведення аудиту даних і збереження ролі людського редакторського контролю.

Ключові слова: нові медіа, алгоритмічний контроль, штучний інтелект, великі дані, алгоритмічна цензура, інформаційні сліпі зони, AI Curtain, журналістика даних, медіа, медіа пропаганда, журналістика, масові комунікації.

Отримано редакцією видання / Received: 11.12.2025
Прорецензовано / Revised: 17.12.2025
Схвалено до друку / Accepted: 22.12.2025
Опубліковано на сайті видання / Published on the publication's website: 30.01.2026