

AI Agents and the Public Sphere: The Collapse of Democratic Discourse in the Age of the "Dead Internet"

Darijo Čerepinko

Department of Communication, Media and Journalism, University North, Koprivnica, Croatia

ABSTRACT

This article explores the decline of the public sphere amid automated content generation, algorithmic mediation, and AI agents. Drawing on Habermas and its critiques, it examines how digital platforms have undermined the conditions for democratic discourse. Automated bots now generate over half of all internet traffic, with malicious bots accounting for 37%, while Section 230 protections shield platforms from liability, blurring the line between genuine opinion and manufactured consensus. The article argues that the traditional link between capitalism and liberal democracy has been disrupted, and, as seen in China and some other successful global economic powers, economic growth can occur without liberal institutions. Big tech firms, operating under these dynamics, have little incentive to uphold democratic functions, and without substantial regulatory intervention, the collapse of the public sphere threatens the foundations of Western liberal governance.

Key words: public sphere, artificial intelligence, democracy, communicative capitalism

Introduction

The public sphere, as conceptualized by Jürgen Habermas¹, represents a critical space between the private domain of individual life and the coercive apparatus of the state – a realm where citizens engage in rational-critical debate, form public opinion, and thereby influence political authority. This intermediary space has historically served as the foundation for democratic legitimacy in liberal societies, providing a mechanism through which private individuals could negotiate the terms of their relationship with power, determining which rights they would surrender to enable functional social organization and which they would retain as inviolable aspects of human dignity.

However, the contemporary public sphere faces an existential crisis of unprecedented magnitude. Recent empirical research reveals a startling transformation in the composition of online discourse: automated bots now account for more than half of all internet traffic, with malicious bots alone comprising 37% of global web activity². This finding represents more than a mere technical curiosity; it signals a fundamental rupture in the epistemological foundations of democratic communication. When the majority of content circulating through the primary

channels of public discourse originates from non-human actors, the very concept of "public opinion" becomes problematic, if not meaningless.

The 2025 Imperva Bad Bot Report² documents that automated bot traffic surpassed human-generated traffic for the first time in 2024, constituting 51% of all web traffic. This shift is largely attributed to the proliferation of generative artificial intelligence and Large Language Models (LLMs), which have dramatically simplified the creation and deployment of sophisticated bots. Bad bots – those created with malicious intent – increased from 32% of internet traffic in 2023 to 37% in 2024, marking the sixth consecutive year of growth in malicious automated activity. The sophistication of these bots has evolved considerably; they now employ advanced techniques such as residential IP addresses, virtual private networks, browser identity spoofing, and increasingly effective CAPTCHA circumvention through AI-enabled systems. Once dismissed as a conspiracy in the 2010s, the "Dead Internet theory" – which claimed that most online content and communication is generated by bots for bots and was largely discussed in obscure corners of the internet such as 4chan – now seems increasingly plausible.

This transformation occurs within a specific legal and regulatory framework that has profound implications for the functioning of the public sphere. In the United States, Section 230 of the Communications Decency Act³ shields technology platforms from liability for user-generated content, creating an asymmetric accountability structure wherein platforms profit from engagement-maximizing algorithmic curation while bearing minimal responsibility for the societal consequences of that curation. This legal architecture, combined with the economic incentives of the attention economy, has created conditions wherein the degradation of democratic discourse may actually enhance platform profitability.

The situation is further complicated by the concentration of information distribution in the hands of a small number of technology corporations. Research by Barberi-Bartov et al.⁴ demonstrates that fewer than 2,000 users were responsible for 80% of misinformation shared during the 2020 U.S. presidential election campaign. Similarly, the Center for Countering Digital Hate⁵ found that just twelve individuals generated approximately 65% of all vaccine-related misinformation across major social media platforms during a critical period of the COVID-19 pandemic. These findings suggest that the distortion of public discourse is not merely a consequence of decentralized network effects but results from concentrated interventions by specific actors who exploit the structural vulnerabilities of digital communication systems.

The traditional media ecology⁶, which once provided at least partial mechanisms for quality control, fact-checking, and editorial oversight, has been displaced by algorithmic curation systems optimized for engagement rather than accuracy or democratic value. Citizens no longer consume information from a limited set of professional journalistic sources that, despite their limitations and biases, operated under at least nominal commitments to truth and professional ethics. Instead, information consumption is increasingly mediated by algorithms that prioritize content based on its capacity to generate emotional responses and sustain attention, regardless of its veracity or social utility.

Prior to the rise of social media and artificial intelligence, mass media served as a crucial component of the public sphere – a space where individuals could encounter diverse perspectives, access information about social issues, and develop shared understandings of social reality. Citizens had opportunities to engage with news, validate their perceptions, and assign meaning to events through processes of social construction that, while imperfect, at least maintained some connection to consensual reality. Now, however, the social cues that enable such collective meaning-making have become fundamentally blurred and distorted. This article examines these transformations through multiple analytical lenses.

The Public Sphere and Democracy

Jürgen Habermas's concept of the public sphere, articulated most comprehensively in *The Structural Transformation of the Public Sphere*⁷, identifies a historically specific formation that emerged in eighteenth-century Europe as bourgeois society developed institutions and practices of rational-critical debate. This sphere constituted "a realm of our social life in which something approaching public opinion can be formed"⁸ – a space analytically distinct from both the intimate domain of family life and the coercive apparatus of state power.

The emergence of the public sphere represented a fundamental rupture with feudal modes of political representation. In the feudal order, the public manifested as the representation of power itself: monarchs, nobility, and ecclesiastical authorities "represented their power before the people instead of representing the people". Sovereignty was displayed rather than deliberated; authority was performed rather than justified through argumentation. The public sphere, by contrast, established a radically different principle: the legitimacy of political authority would henceforth depend upon its capacity to withstand rational scrutiny and critical examination by private citizens acting in their capacity as members of the public.

This transformation involved several interconnected developments. First, the expansion of market economies created a class of property-owning individuals whose economic interests required predictable legal frameworks and rational administration rather than arbitrary sovereign will. Second, the development of print enabled the circulation of ideas beyond local communities, creating the possibility of a reading public that could engage with political questions at a distance from immediate power relations. Third, the emergence of coffee houses, salons, and voluntary associations provided physical spaces where individuals could gather for discussion freed from the constraints of courtly etiquette and aristocratic hierarchy.

Crucially, the public sphere presupposed equality among participants, at least in principle. Arguments were to be evaluated based on their rational merit rather than the social status of their proponents. As Habermas notes, this represented an ideal that was never fully realized in practice, but it nevertheless established a normative standard against which actual practices could be criticized. The public sphere served as "a network for communicating information and points of view; the streams of communication are filtered and synthesized in such a way that they coalesce into public opinions"⁹.

Institutional expressions of the public sphere included journalism, debate societies, literary clubs, and political associations. These institutions operated according to specific norms: publicity (arguments must be made openly rather than secretly), rationality (claims must be supported by reasons rather than appeals to tradition or authority), and inclusivity (in principle, participation was open to all capable of rational discourse). Most importantly, "public opinion was not a mere aggregate of private opinions, but the result of a process of rational-critical debate"¹⁰. This distinction is crucial: public opinion, properly con-

ceived, emerges through deliberation and argument rather than through the simple aggregation of pre-existing preferences.

Following Immanuel Kant, whom Habermas identifies as "the first to formulate publicity as the bridge between morality and politics", the public sphere serves not merely as a space for the expression of opinion but as a mechanism for the legitimation of political decisions. Power exercised in accordance with publicly debated principles possesses a normative validity that coercive force alone cannot provide. This Kantian insight establishes the public sphere as essential to the moral foundation of political authority in liberal democratic societies.

The public sphere, in this classical formulation, performs several critical functions for democratic governance. First, it serves as an information distribution mechanism, ensuring that citizens have access to knowledge about political affairs. Second, it provides a deliberative space where diverse perspectives can be articulated and subjected to critical examination. Third, it generates public opinion through processes of argumentation rather than mere preference aggregation. Fourth, it holds political authority accountable by subjecting governmental actions to ongoing public scrutiny. Fifth, it legitimates political power by ensuring that decisions can withstand rational examination by those subject to them.

This theoretical framework helps clarify what is at stake in the contemporary transformation of communication systems. The public sphere is not merely a desirable feature of democratic societies but a functional prerequisite for their operation. Without mechanisms for rational-critical debate, public opinion formation, and the accountability of power to public scrutiny, the distinction between democratic governance and other forms of political organization collapses. A society may retain the formal apparatus of democracy – elections, parliaments, constitutional protections, etc. – while losing the substantive conditions that make these institutions meaningful vehicles for collective self-governance.

The relationship between the public sphere and democracy is thus not contingent but constitutive. Democratic legitimacy rests upon the principle that political authority derives from the consent of the governed, but meaningful consent requires conditions under which citizens can form informed judgments about political questions, deliberate collectively about shared concerns, and hold power accountable to public reason. The public sphere provides precisely these conditions. Its degradation therefore represents not a peripheral problem but a direct threat to the foundations of democratic governance.

Critiques of the Public Sphere

While Habermas's conceptualization of the public sphere has profoundly influenced democratic theory, it has also generated substantial critical engagement. These critiques do not merely identify empirical deviations from the

ideal but challenge the theoretical premises and normative assumptions underlying the Habermasian framework. Understanding these critiques is essential for grasping both the limitations of public sphere theory and the complexity of contemporary challenges to democratic discourse.

Habermas himself, in later work, acknowledged significant problems with the classical public sphere. The transformation "from a culture-debating public to a culture-consuming public" fundamentally altered the character of public discourse. The public sphere "once again becomes a platform for displaying power before the people rather than allowing them to participate in it". This refeudalization of the public sphere parallels in some respects the original feudal mode of representation, albeit through different mechanisms. Rather than monarchical spectacle, contemporary power displays itself through public relations, advertising, and the manipulation of images.

The transformation of the liberal constitutional state into a social-welfare state changed the structural relationship between citizens and political authority. A critical public engaging in rational debate was increasingly replaced by "a public of consumers of public services". Citizens no longer primarily participate in political will-formation through deliberation but "react to administrative decisions as consumers react to commodities". This consumerist orientation fundamentally alters the character of political engagement, substituting service delivery evaluation for collective deliberation about shared purposes.

Moreover, the spontaneous formation of public opinion through critical debate has been increasingly displaced by its manufacture through strategic communication. As Habermas notes, "public opinion has been manufactured" through public relations techniques, opinion polling that shapes the opinions it purports to measure, and mass media that circulate pre-packaged interpretations rather than facilitating critical dialogue. Nevertheless, Habermas maintains that "the public sphere can survive only if the mass of the population learns to act as a public, to use the medium of communication rationally and critically", suggesting that media literacy and critical engagement remain possible responses to these degradations.

However, more fundamental critiques question whether the idealized public sphere ever existed or could exist. Holub argues that "Habermas's public sphere never truly existed; it is a regulative idea projected backward into history"⁷. This critique suggests that Habermas conflates a normative ideal with historical description, obscuring the extent to which the actual bourgeois public sphere was always already compromised by exclusion, power, and strategic interest. The danger of such idealization is that it obscures the necessity of more radical transformations by holding up a mythical golden age as a model for restoration.

Holub further challenges the epistemological assumptions underlying Habermas's theory of communicative rationality: "Habermas assumes that communication,

once freed from distortion, will naturally converge on truth. This assumption neglects the opacity, irony, and indeterminacy inherent in language itself". From this perspective, the ideal speech situation, a hypothetical condition of perfect communication free from domination and distortion, presupposes a transparency of meaning and convergence on truth that the nature of language itself renders impossible. Communication is always already mediated by linguistic structures that carry historical sedimentation, cultural specificity, and irreducible ambiguity.

Moreover, the ideal speech situation may itself encode particular power relations. As Holub observes, it "presupposes a form of discourse already saturated with power – the power of Western reason to define what counts as intelligible". The norms of rational argumentation, far from being universal and neutral, reflect specific cultural and historical formations. Privileging certain forms of speech (rational, deliberative, argument-based) over others (emotional, testimonial, performative) may systematically disadvantage groups whose communicative practices have developed under different conditions.

Michel Foucault's work provides a more radical challenge to Habermasian assumptions about the relationship between reason and power. Rather than treating power as an external constraint on reason that can be overcome through ideal communicative conditions, Foucault argues that "power does not stand outside reason; it defines what can count as reasonable"⁸. From this perspective, the very categories through which we think and argue are themselves products of power relations. The aspiration to transcend power through rational communication may therefore be fundamentally misconceived.

Nancy Fraser's feminist critique highlights the exclusions built into the bourgeois public sphere from its inception: "The bourgeois public sphere was premised on the exclusion of those whose material conditions prevented them from participating as equals"⁹. Women, workers, racial minorities, and colonized peoples were systematically excluded from participation, not as accidental failures of implementation but as constitutive conditions of the public sphere's operation. The idealized space of rational debate among equals was possible only because it rested upon a foundation of exclusion and domination rendered invisible by the very structure of public discourse.

Furthermore, the assumption of a singular public sphere obscures the existence of multiple, competing publics. As Crossley and Roberts note, "there was never one public sphere but a plurality of counter-publics that challenged its hegemony"¹⁰. Subaltern groups developed their own spaces for communication and identity formation, operating according to different norms and serving different purposes than the bourgeois public sphere. Recognizing this multiplicity challenges the implicit universalism of Habermasian theory while opening possibilities for more genuinely inclusive democratic practice.

Pierre Bourdieu's concept of cultural capital provides another critical lens. "The public sphere is not a neutral

arena but a field structured by the distribution of capital and habitus"¹⁰. Participation in public discourse requires specific competencies (linguistic facility, educational credentials, familiarity with cultural references...) that are unequally distributed according to class position. The apparent equality of rational argumentation masks deeper structures of inequality that determine who can effectively participate. As Crossley and Roberts observe, "rational communication presupposes a distribution of embodied competence; not everyone has an equal right to appear or to be heard"¹⁰.

More recent democratic theory has emphasized the limitations of purely rational deliberation and the importance of affect, passion, and contestation. As Roberts argues, "democracy begins where the critique of democracy becomes possible"¹¹, suggesting that the health of democratic systems depends not on achieving consensus through rational deliberation but on maintaining spaces for fundamental disagreement and contestation. Benhabib similarly argues that "the public sphere must be porous to difference; otherwise its universality becomes domination"¹².

Young emphasizes the need for more expansive conceptions of democratic communication: "Inclusion means making room for forms of expression that do not conform to dominant norms of speech"¹³. Testimony, storytelling, emotional expression, and performative action may all constitute legitimate forms of political communication, challenging the privileged status of rational argumentation in traditional public sphere theory. Mouffe goes further, arguing that "without passion, there is no democratic engagement; without disagreement, no politics"¹⁴. Democracy requires not the transcendence of conflict through rational consensus but the productive channeling of agonistic contestation.

The relationship between emotion and reason in public discourse has been particularly reevaluated. Jaggar argues that "emotions are not irrational; they are the means through which values become embodied and shared"¹⁵. From this perspective, the exclusion of emotion from legitimate public discourse represents not a purification of reason but an impoverishment of democratic communication. Ahmed extends this insight by observing that "anger can be a mode of public reason when it exposes injustice"¹⁶. Affective responses to oppression and inequality may provide critical insights unavailable to dispassionate analysis.

These critiques do not render the concept of the public sphere obsolete but demand its substantial reconstruction. A democratically adequate public sphere must acknowledge rather than obscure power relations, embrace rather than suppress difference, recognize multiple rather than singular publics, and accommodate diverse rather than uniform modes of expression. The normative ideal remains important but this ideal must be pursued with greater attention to the material, cultural, and affective conditions that enable or prevent meaningful participation.

Evolution of the Public Sphere in the Digital World

The emergence of the internet and, subsequently, social media platforms has fundamentally transformed the structural conditions of the public sphere. These transformations cannot be understood as merely technical changes in communication infrastructure but represent qualitative shifts in the nature of public discourse, information distribution and political engagement. However, scholarly assessments of these changes have evolved considerably, moving from initial optimism about democratization through digital communication to increasing concern about the erosion of the epistemic and institutional foundations required for democratic deliberation.

Early internet utopianism emphasized the potential for digital communication to realize Habermasian ideals more fully than previous media. The internet appeared to overcome many of the exclusions and limitations of traditional mass media: it lowered barriers to entry for political speech, enabled direct participation rather than passive consumption, facilitated communication across geographical and social boundaries, and created spaces for deliberation relatively insulated from state and corporate control. The proliferation of blogs, forums, and early social media platforms seemed to promise a revitalization of the public sphere through the democratization of communication.

However, Dean¹⁷ offered an early critical assessment, arguing that the internet does not fit the norms of Habermas's model "not because something is missing, but because the very idea of the public sphere no longer matches the historical conditions." According to Dean, appeals to "the public" in digital contexts often serve to legitimize market regulations and the concentration of power rather than enabling democratic participation. The ideology of publicity creates an impression of democratic transparency while actually reinforcing economic inequalities and contributing to depoliticization. The multiplication of voices and the ease of participation do not necessarily translate into enhanced democratic capacity; instead, they may produce what Dean calls "communicative capitalism" – a system where communication itself becomes capital, measured in clicks, likes, and shares, with minimal connection to political efficacy or social transformation.

The rise of social media platforms intensified these contradictions. Vaccari and Valeriani¹⁸ document how "the rise of social media has profoundly altered how citizens access, process, and share political information". Political participation now includes "expressive and connective actions that take place on digital platforms", expanding beyond traditional forms of civic engagement like voting, campaigning, and attending public meetings. Moreover, "politics is no longer confined to institutions; it permeates the routines and interactions of everyday digital life". This expansion might appear democratizing, but it also dissolves the boundaries between political and non-political domains in ways that may undermine rather than enhance democratic deliberation. Critically, while "social

media can lower the costs of participation, they do not eliminate barriers of knowledge and motivation". The reduction of technical and economic barriers to political expression does not automatically translate into more informed or effective political engagement. Information abundance may paradoxically lead to information poverty when individuals lack the time, skills, or motivation to critically evaluate and synthesize available information. The problem is not scarcity of information but its overwhelming superabundance combined with increasingly sophisticated techniques for manipulation and deception.

The structural characteristics of digital public discourse differ fundamentally from those of the classical public sphere. "The public sphere today is not a place of gathering but a network of flows: of information, images and affects"¹⁹. The spatial metaphor of the "sphere", suggesting a bounded, relatively stable domain, fails to capture the fluid, distributed, and perpetually reconfiguring nature of digital communication. Furthermore, "publicness is no longer simply spoken; it is programmed, streamed, and filtered". Algorithms rather than editorial judgment determine visibility; computational processes rather than human deliberation structure information access.

Dean¹⁹ developed her analysis of communicative capitalism to address these transformations. In contemporary society, she argues, communications themselves become capital. Likes, clicks, shares, and other engagement metrics serve as measures of value that can be monetized through advertising and data extraction. This transformation fundamentally alters the relationship between communication and democratic practice. In classical public sphere theory, communication serves as a means toward the formation of public opinion and collective will. In communicative capitalism, communication becomes the production of ever-increasing volumes of content that generate engagement regardless of its truth value or political significance.

This shift has profound implications for the possibility of rational-critical discourse. When platforms optimize for engagement rather than truth or democratic value, content that provokes emotional responses (outrage, fear, anxiety...) gains algorithmic advantage over content that promotes careful deliberation. Research by Vosoughi, Roy, and Aral²⁰ demonstrates that false news spread farther, faster, and deeper than true news on Twitter. False stories reached up to 19 "degrees" of separation and were shared 6-10 times more rapidly than true stories, which typically reached only 10 degrees. Significantly, this differential spread was driven primarily by human users rather than bots, suggesting that false information possesses structural advantages in social media ecosystems due to its novelty, emotional engagement, and perceived informational value.

The displacement of traditional journalism by social media as a primary information source has further undermined the institutional infrastructure of the public sphere. Jamieson and Cappella²¹ and Tsfati and Ariely²² found that use of the internet as a primary source of information is

associated with reduced trust in institutions. Similarly, Brants²³, Ladd²⁴, and Allcott and Gentzkow²⁵ documented declining trust in traditional media, with fake news and low media illiteracy contributing to uncritical belief in content regardless of its veracity. The Reuters Institute's 2017 Digital News Report²⁶ showed a key turning point as, for the first time, most people got their daily news online rather than from traditional media. This transition is not merely a change in medium but a transformation in the institutional ecology of information production and distribution. Traditional journalism, despite its many limitations and biases, operated under professional norms that at least nominally prioritized accuracy, verification, and editorial oversight. These norms were enforced through professional training, organizational culture, legal liability, and reputational mechanisms. Digital platforms, by contrast, disclaim responsibility for content accuracy, optimization algorithms that may actively promote misinformation if it generates engagement, and operate under legal protections that insulate them from liability for user-generated content.

The transformation of information access patterns has occurred with remarkable rapidity. Within a single generation, the dominant mode of encountering news and political information has shifted from exposure to professionally produced content through mass media channels to personalized streams of mixed professional and amateur content curated by algorithmic systems optimized for engagement. This transformation has occurred without corresponding development of critical literacies, institutional safeguards, or regulatory frameworks adequate to the new environment.

Moreover, the architecture of social media platforms shapes discourse in ways that undermine conditions necessary for democratic deliberation. The affordances of these platforms like limited character counts, emphasis on rapid response, algorithmic amplification of viral content or design features that discourage extended argumentation systematically privilege certain forms of communication over others. The public sphere requires spaces for sustained, reasoned argumentation; social media platforms provide environments optimized for brief, emotionally charged reactions.

The result is what might be called a structural disjunction between the technical capabilities of digital communication systems and the normative requirements of democratic discourse. The technology enables unprecedented connectivity and information distribution, but the institutional forms, economic incentives, and design choices shaping its deployment actively undermine the conditions necessary for the formation of informed public opinion through rational-critical debate.

Problems of the Digital World and the Ideal Speech Situation

The ideal speech situation, a central concept in Habermas's theory of communicative action, presupposes condi-

tions of free, equal, and undistorted communication where participants can raise claims to truth, rightness, and sincerity without coercion or strategic manipulation. This ideal provides a normative standard for evaluating actual communicative practices and identifying systematic distortions. The contemporary digital public sphere deviates from this ideal in ways that fundamentally compromise its democratic capacity.

Misinformation superspreaders and concentrated manipulation

Research has consistently demonstrated that misinformation distribution follows a highly skewed pattern, with small numbers of actors responsible for disproportionate shares of false information dissemination. The study by Barbieri-Bartov et al.⁴ found that fewer than 2,000 users—the "supersharers"—were responsible for 80% of misinformation shared during the 2020 election campaign. These supersharers had an average age of approximately 58 years, were predominantly older white women, majority Republican, and primarily located in "red" states such as Florida, Texas, and Arizona. Importantly, their activity did not appear automated but rather involved manual and persistent retweeting, suggesting dedicated effort rather than bot-driven distribution.

Similarly, the Center for Countering Digital Hate⁵ identified twelve individuals who generated approximately 65% of all vaccine-related misinformation on Facebook, Instagram, and Twitter between February 1 and March 16, 2021. Robert F. Kennedy Jr. was identified as the top contributor. This concentration of misinformation production and distribution represents a profound vulnerability in the structure of the digital public sphere. A relatively small number of actors can shape the informational environment experienced by millions, distorting collective understanding of critical issues.

Research on information diffusion mechanisms reveals the crucial role of opinion leaders and social influence in shaping belief formation. Katz and Lazarsfeld²⁷ demonstrated that media messages are often first received by "opinion leaders" who then interpret and pass information to others in their social networks. Zaller²⁸ extended this insight, showing that individuals receive information from elites, accept it based on predispositions, and sample from this information when forming opinions. Nicholson²⁹ found that while elites can shape opinions, cues related to group membership often dominate elite cues in opinion formation.

In digital environments, these influence mechanisms operate with enhanced potency. Moussaid et al.³⁰ demonstrated that individuals' opinions can converge or diverge based on peer influence, with social media platforms providing unprecedented opportunities for such influence at scale. Walther³¹ showed that individuals assess information credibility based on cues that are less controllable by the source, often relying on external indicators like social proof rather than content evaluation. Naab et al.³² found that critical user comments can reduce perceived credibil-

ity of news articles, though this effect depends on whether the critical comment receives likes and can be partially counteracted by reply comments containing counter-speech.

These dynamics create conditions where misinformation can spread rapidly through social networks while corrections struggle to achieve comparable reach. The asymmetry between the viral spread of false information and the limited circulation of corrections represents a structural problem for democratic discourse.

The experience of second-hand politics

The transformation of political experience through digital mediation has profound implications for democratic engagement. As Vaccari and Valeriani¹⁸ observe, "most people experience politics second-hand, through the voices of their peers and influencers". This mediation is not new, mass media have always shaped political perception to some extent, but the nature of mediation has changed fundamentally. Traditional journalism provided at least some professional gatekeeping and fact-checking, however imperfect. Social media provides no such filtering; instead, information reaches users through networks of peers whose credibility is assessed through social rather than professional criteria.

This shift has several consequences. First, the credibility of information becomes detached from its veracity and instead depends on its source within social networks. Users are more likely to believe and share information from trusted connections even when that information is false, creating echo chambers where misinformation circulates freely within ideologically homogeneous networks. Second, the volume and velocity of information flow overwhelm individual capacity for critical evaluation. Users make rapid judgments based on heuristics and affective responses rather than careful assessment of evidence. Third, the intermingling of political and personal content on social media platforms blurs boundaries between different domains of life, potentially intensifying political polarization while also trivializing political questions.

The problem is exacerbated by platform design features that prioritize engagement over accuracy. Algorithms optimize for content that generates reactions creating incentives for emotionally charged, controversial, or sensationalistic content regardless of its truth value. This optimization process systematically distorts the informational environment, making highly engaging false information more visible than less engaging true information. Users encounter information that platforms have determined will maximize their engagement rather than information that will enable informed democratic participation.

Algorithmic control of visibility

Perhaps the most fundamental challenge to the ideal speech situation in digital environments concerns the algorithmic mediation of visibility. As Vaccari and Valeri-

ani¹⁸ notice, "algorithms decide who is seen and who remains invisible". This algorithmic gatekeeping operates opaquely, according to criteria determined by platform companies based on business objectives rather than democratic values. The result is that "democracy cannot thrive when its visibility is governed by private code".

The opacity of algorithmic curation systems creates several problems. First, users cannot know why they see particular content or why and what other content is hidden from them, making it impossible to assess whether their social media feeds provide an adequate basis for democratic participation. Second, the criteria determining visibility are proprietary and can be changed unilaterally by platforms without public accountability. Third, these systems can be and are manipulated by actors who understand their mechanics, creating advantages for sophisticated operators over ordinary users. These encoded priorities shape the structure of public discourse without explicit acknowledgment or democratic legitimation. As Vaccari and Valeriani¹⁸ argue, "algorithmic design should be treated as a matter of public policy, not private innovation".

The concentration of algorithmic power in the hands of a small number of technology corporations represents a profound challenge to democratic governance. These corporations exercise unprecedented influence over the formation of public opinion, not through direct control of content but through their control over visibility and distribution. This influence operates at a level below conscious awareness for most users, who experience algorithmically curated feeds as neutral or natural rather than as products of specific corporate choices optimized for commercial rather than democratic objectives.

The bot problem and synthetic public opinion

The most radical challenge to the integrity of the digital public sphere emerges from the proliferation of automated agents – bots – that generate content, simulate human interaction, and shape the informational environment. As noted in the introduction, recent data indicates that automated bots now account for more than 51% of all internet traffic, with malicious bots alone comprising 37% of global web activity². The implications of this development are profound. If public opinion formation depends upon processes of communication among citizens, what does it mean when the majority of "participants" in these processes are not citizens at all but automated programs designed to influence, manipulate, or simulate human interaction? The concept of "public opinion" presupposes a public, a collectivity of human beings capable of forming judgments based on reasons, experiences, and values. When this presupposition can no longer be taken for granted, the epistemological foundations of democratic discourse collapse.

The sophistication of contemporary bots has evolved considerably. Modern bots employ advanced techniques including residential IP addresses to mask their automated nature, virtual private networks to obscure their ori-

gins, browser identity spoofing to mimic human users, and increasingly effective CAPTCHA circumvention through AI-enabled systems. The integration of Large Language Models has dramatically enhanced bots' capacity to generate contextually appropriate, linguistically sophisticated content that is increasingly difficult to distinguish from human-generated text.

This technological evolution creates what might be called the authenticity crisis of the digital public sphere. Users encountering content can no longer reliably determine whether they are engaging with authentic human expression or algorithmically generated simulation. This uncertainty corrodes the trust necessary for democratic discourse. When citizens cannot distinguish authentic public opinion from manufactured consensus, the very possibility of collective will-formation through deliberation becomes questionable.

The use of bots for political manipulation has been extensively documented³³. Automated accounts amplify particular messages, create false impressions of grassroots support, harass political opponents, spread disinformation, and manipulate trending topics. These interventions distort the informational environment in ways that systematically undermine informed democratic participation. Citizens form opinions based on what they perceive as widespread public sentiment, unaware that this "sentiment" is partially or largely manufactured by automated systems.

The problem extends beyond overtly political manipulation. Commercial bots that inflate engagement metrics, generate fake reviews, or manipulate search rankings contribute to a generalized degradation of information integrity. When users learn that significant portions of online activity are fake, they become skeptical of all online information, including authentic content. This skepticism, while perhaps rational at the individual level, is corrosive to the collective capacity for truth-seeking that democratic deliberation requires.

The cumulative effect of these developments is the erosion of what might be called the epistemic commons: the shared body of knowledge, facts, and frameworks that enable collective deliberation. Democratic discourse presupposes that participants share at least some common understanding of reality, even as they disagree about values, priorities, and policies. When this common ground dissolves, reasoned deliberation becomes impossible. This fragmentation, however, is not accidental but results from the structural features of digital communication platforms combined with human psychological tendencies.

Problems with Media Systems

The structural vulnerabilities of the digital public sphere cannot be understood apart from the legal and regulatory frameworks governing communication platforms. These frameworks shape platform incentives, determine accountability structures, and establish the boundaries of permissible activity. In the United States, Section 230 of

the Communications Decency Act has played a particularly significant role in enabling the current dysfunctional state of digital discourse.

Section 230 and platform immunity

Section 230 of the Communications Decency Act provides that "no provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider"³⁴. This provision shields online platforms from liability for user-generated content, establishing a fundamental asymmetry: platforms can profit from hosting and distributing content without bearing responsibility for its accuracy, legality, or social consequences.

The original justification for Section 230 protection was to enable the development of online platforms without exposing them to ruinous liability for the vast volumes of user content they could not feasibly monitor. This justification may have been reasonable in the early days of the internet, when platforms operated as relatively passive hosts of content. However, contemporary platforms do not merely host content; they actively curate, recommend, and amplify it through algorithmic systems optimized for engagement. The distinction between passive hosting and active curation problematizes the continued applicability of Section 230 protection.

The immunity provided by Section 230 creates a moral hazard. Platforms face minimal consequences for algorithmic choices that promote misinformation, extremism, or other socially harmful content, provided this content is user-generated. Indeed, to the extent that such content generates engagement and thereby increases advertising revenue, platforms have positive incentives to facilitate its spread. This incentive structure systematically privileges commercial objectives over democratic values, epistemic integrity, or social cohesion.

The contrast with traditional media is instructive. Newspapers, television networks, and other mass media operate under legal frameworks that impose liability for defamation, incitement, and other harmful content. These liability structures, while imperfect, create incentives for editorial oversight and fact-checking. Digital platforms face no comparable incentives.

Defenders of Section 230 argue that removing or limiting platform immunity would lead to excessive censorship, as platforms would suppress potentially problematic content to avoid liability risk. Nevertheless, traditional mass media function within this legal framework, facilitating free expression and rational-critical debate, while enjoying extensive protection of the "First Amendment"³⁴ through legal rulings and established precedents.

The legal implications of the First Amendment for social media regulation are highly complex and lie beyond the scope of this article. Nevertheless, no political system has failed to regulate critical communication infrastructure. In liberal democracies, such regulation considers fundamental rights and freedoms, whereas authoritarian

regimes impose stricter and less transparent controls. Digital communication networks, particularly in the United States, where most are headquartered, remain largely unregulated. Accordingly, platforms such as TikTok and Chinese social networks are subject to far more extensive oversight in their countries of origin than in their Western counterparts.

Media system models and regulatory approaches

Hallin and Mancini³⁵ identified the North Atlantic (Liberal) Model of media systems, characterized by professional journalism with relatively strong insulation from political control, market orientation, and quality control through mutual competition. This model presupposes that market mechanisms, combined with professional norms and competition, will produce information of adequate quality for democratic purposes. The transformation of the media landscape has called this presupposition into question.

Brüggemann et al.³⁶ described the Western system as a regulated-market type, defined by limited support for public-service media and few opportunities to obtain public funding for communication services, with strict regulation of ownership but minimal content regulation. This approach reflects the liberal commitment to free speech and skepticism toward government involvement in information provision. However, the adequacy of this regulatory approach in the digital age is increasingly doubtful.

The market-based model assumes that competition will discipline producers and that consumer choice will drive quality. In digital media markets, these assumptions fail. Network effects, data advantages, and algorithmic manipulation create winner-take-all dynamics that undermine competition. Users lack the information and expertise to effectively evaluate content quality, particularly when sophisticated misinformation techniques are employed. The business model of advertising-supported platforms creates incentives for engagement maximization rather than information quality, systematically distorting content toward the sensational, emotional, and false.

Alternative regulatory models deserve consideration. Public service broadcasting models, common in European democracies, provide mechanisms for funding quality journalism independent of both commercial and direct governmental control. Regulatory frameworks that impose duties of care on platforms could address some dysfunctions without requiring direct content regulation. Transparency requirements could make algorithmic curation systems more accountable without determining their specific operation. Competition policy could address the concentration of platform power that enables current abuses.

The challenge is designing regulatory interventions that address genuine problems without creating new ones. Overly restrictive regulation could suppress legitimate speech, create barriers to entry that protect incumbent platforms, or enable government control over information in ways incompatible with democratic values. However, the current approach of minimal regulation combined

with broad platform immunity has produced a digital public sphere profoundly dysfunctional for democratic purposes. The question is not whether to regulate but how to regulate effectively and appropriately.

Conclusion

Capitalism without democracy and possible responses to the power imbalance

The fundamental argument of this article is that the contemporary transformation of the public sphere represents not a correctable malfunction but a structural condition reflecting deeper tensions in the relationship between capitalism and democracy. The classical liberal assumption that these systems are mutually reinforcing, that market economies require democratic governance and democratic governance requires market economies, no longer holds, if it ever did. Authoritarian capitalist models, exemplified by, for example, China, Singapore or Middle-eastern theocracies demonstrate that rapid economic growth, technological innovation, and rising living standards can be achieved without institutions of western type democracy, civil liberties, or a normatively functioning public sphere. Over the past four decades, China's economic trajectory stands as perhaps the most striking example of capitalism operating without the foundational tenets of liberal democracy, following Deng Xiaoping's pragmatic observation that "it doesn't matter if a cat is black or yellow, as long as it catches mice"³⁷.

These examples are not merely curiosities but represent an alternative pathway to modernity that directly challenges liberal democratic assumptions. If capitalism can thrive without democracy, then capitalist actors, particularly large tech corporations, oligarchs and politicians with authoritarian tendencies, have no necessary interest in preserving democratic institutions.

The concentration of digital communication infrastructure in the hands of a few technology corporations poses a significant challenge to democratic governance, as these platforms influence public discourse while largely escaping accountability. Addressing this requires a combination of regulatory, institutional, technical, and cultural interventions. Regulatory measures might include limiting platform immunity, enforcing transparency, promoting interoperability, applying antitrust laws, strengthening privacy protections, imposing public-interest obligations, and enabling user participation in governance. Institutional alternatives could involve public service, cooperative, or nonprofit platforms, as well as investment in journalism and digital literacy to support informed democratic deliberation. Technical solutions such as algorithmic reform, verification tools, and community-based moderation can mitigate manipulation, while cultural and political mobilization—including citizen advocacy, labor organizing, and international cooperation—is essential to

ensure sustained accountability and democratic oversight of powerful digital platforms.

This analysis suggests a pessimistic conclusion: the public sphere may not be saved because powerful actors do not wish to save it. The classical liberal assumption that enlightened self-interest would lead elites to support democratic institutions presupposed that their own security and prosperity depended upon democratic stability. This presupposition no longer holds. However, this pessimistic conclusion must be tempered by recognition of contingency and agency. The trajectory described in this article is not inevitable but results from specific choices, policies, and institutional arrangements that could be different. There are constituencies with interests in preserving democratic institutions and a functional public

sphere. Citizens who value self-governance, journalists who depend on informed publics, civil society organizations that require spaces for mobilization, and even some business actors who recognize that long-term stability and legitimacy require democratic foundations might constitute a coalition capable of demanding reform. The question is whether such a coalition can be organized and sustained against powerful interests that benefit from the current arrangement. It requires sustained political mobilization capable of overcoming the resistance of powerful actors who benefit from the status quo. Addressing this challenge requires building or rebuilding institutions, practices, and forms of consciousness capable of critical distance from the dominant systems of communication and information distribution. Whether such rebuilding is possible remains an open question.

REFERENCES

1. HABERMAS J, *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society* (MIT Press, Cambridge, 1989).
2. IMPERVA, 2025 *Imperva Bad Bot Report*. <https://www.imperva.com/resources/resource-library/reports/bad-bot-report/>.
3. 47 U.S.C. § 230 (Communications Decency Act).
4. BARBIERI BARTOV A, et al., *Science*, 381 (2024) 392.
5. CENTER FOR COUNTERING DIGITAL HATE, *The Disinformation Dozen*. <https://www.counterhate.com/disinformationdozen>.
6. MCQUAIL D, DEUZEM, *McQuail's Media and Mass Communication Theory* (SAGE Publications, London, 2020).
7. HOLUB RC, *Jürgen Habermas: Critic in the Public Sphere* (Routledge, London, 1991).
8. FOUCAULT M, *Power/Knowledge: Selected Interviews and Other Writings, 1972-1977* (Pantheon Books, New York, 1980).
9. FRASER N, In: CALHOUN C (Ed) *Habermas and the Public Sphere* (MIT Press, Cambridge, 1992) 109.
10. CROSSLEY N, ROBERTS JM (Eds), *After Habermas: New Perspectives on the Public Sphere* (Blackwell Publishing, Oxford, 2004).
11. ROBERTS JM, *Brit J Polit Int Rel*, 6 (2004) 471.
12. BENHABIB S, *Democracy and Difference: Contesting the Boundaries of the Political* (Princeton University Press, Princeton, 1996).
13. YOUNG IM, *Inclusion and Democracy* (Oxford University Press, Oxford, 2000).
14. MOUFFE C, *The Democratic Paradox* (Verso, London, 2000).
15. JAGGAR AM, *Inquiry*, 32 (1989) 151.
16. AHMED S, *The Cultural Politics of Emotion* (Edinburgh University Press, Edinburgh, 2004).
17. DEAN J, *Constellations*, 10 (2003) 95.
18. VACCARI C, VALERIANI A, *Outside the Bubble: Social Media and Political Participation in Western Democracies* (Oxford University Press, Oxford, 2020).
19. DEAN J, *Democracy and Other Neoliberal Fantasies: Communicative Capitalism and Left Politics* (Duke University Press, Durham, 2009).
20. VOSOUGHIS, ROYD, ARALS, *Science*, 359 (2018) 1146.
21. JAMIESON KH, CAPPELLA JN, *Echo Chamber: Rush Limbaugh and the Conservative Media Establishment* (Oxford University Press, Oxford, 2008).
22. TSFATI Y, ARIELY G, *Commun Res*, 41 (2014) 760.
23. BRANTS K, In: PETERS C, BROERSMA M (Eds), *Rethinking Journalism* (Routledge, London, 2013) 15.
24. LADD JM, *Why Americans Hate the Media and How It Matters* (Princeton University Press, Princeton, 2012).
25. ALLCOTT H, GENTZKOW M, *J Econ Perspect*, 31 (2017) 211. doi:10.1257/jep.31.2.211
26. NEWMAN N, *Reuters Institute Digital News Report 2017* (Reuters Institute for the Study of Journalism, Oxford, 2017).
27. KATZ E, LAZARFELD PF, *Personal Influence: The Part Played by People in the Flow of Mass Communications* (Free Press, New York, 1955).
28. ZALLER JR, *The Nature and Origins of Mass Opinion* (Cambridge University Press, Cambridge, 1992).
29. NICHOLSON SP, *J Polit*, 73 (2011) 1165.
30. MOUSSAÏD M, KÄMMER JE, ANALYTIS PP, NETH H, *PLoS ONE*, 8 (2013) e78433. doi:10.1371/journal.pone.0078433
31. WALTHER JB, *J Comput-Mediat Commun*, 14 (2009) 748.
32. NAAB TK, KALCH A, SCHLÜTZ D, *Journalism Stud*, 21 (2020) 749.
33. CHANG HC, et al., In: *Handbook of Computational Social Science, Volume 1* (Routledge, London, 2021) 304.
34. U.S. Const. amend. I.
35. HALLIN DC, MANCINI P, *Comparing Media Systems: Three Models of Media and Politics* (Cambridge University Press, Cambridge, 2004).
36. BRÜGGEMANN M, ENGESSER S, BÜCHEL F, HUMPRECHT E, CASTRO L, *J Commun*, 64 (2014) 1037. doi:10.1111/jcom.12127
37. DENG X, *How to Restore Agricultural Production* (Speech to the Third Plenary Session of the Third Central Committee of the Communist Youth League, 7 July 1962).

D. Čerepinko

Department of Communication, Media and Journalism, University North, Žarka Dolinara Sq. 1, 48000, Koprivnica, Croatia

e-mail: darijo.cerepinko@unin.hr

AI AGENTI I JAVNA SFERA: KOLAPS DEMOKRATSKOG DISKURSA U DOBA „MRTVOG INTERNETA“

SAŽETAK

Ovaj članak istražuje degradaciju javne sfere uslijed automatizirane proizvodnje sadržaja, algoritamskog uređivanja sadržaja i AI agenata. Polazeći od Habermasa i kritika njegovih teorija, autor analizira kako digitalne platforme narušavaju postavke demokratskog diskursa. Automatizirani botovi sada generiraju više od polovice cjelokupnog internetskog prometa, pri čemu zlonamjerni botovi čine 37%, dok pravne odluke sudova SAD-a štite platforme od odgovornosti, zamagljujući granicu između stvarnog javnog mnijenja i umjetno proizvedenog konsenzusa. Članak tvrdi da je tradicionalna poveznica između kapitalizma i liberalne demokracije narušena, te da, kao što se vidi u slučaju Kine i nekih drugih uspješnih globalnih ekonomskih sila, ekonomski rast može biti ostvaren bez demokratsko-liberalnih institucija. Velike tehnološke tvrtke, djelujući unutar ovih tendencija, imaju malo poticaja za očuvanje demokratskih funkcija, a bez značajne regulatorne intervencije, kolaps javne sfere ugrožava temelje zapadnog liberalnog upravljanja.

