BULLETIN
OF "CAROL I" NATIONAL DEFENCE UNIVERSITY

https://buletinul.unap.ro/index.php/en/

Disinformation Dynamics
Unveiling the Impact of Echo Chambers in Shaping Online Public Opinion

Ștefania-Elena STOICA, Ph.D. Student*

*"Carol I" National Defence University, Bucharest, Romania
e-mail: stoica.stefania@myunap.net

Abstract

The proliferation of misinformation and the emergence of echo chambers in the online environment pose significant challenges to modern democracies, directly impacting public opinion and social behaviors. This study focuses on the analysis of a Facebook group centered around a prominent Romanian political figure, boasting 93,800 members and averaging ten daily posts. Using advanced machine learning and AI-based hate speech detection, the study uncovers systematic echo chamber construction and the amplification of misinformation. The findings emphasize the influence of online echo chambers on public opinion and underscore the need to maintain information integrity in the media landscape and communication. This research has important implications for scholars, policymakers, and media practitioners, indicating the critical need to address the challenges posed by misinformation and echo chambers in the online environment.

Keywords:

Background

In the digital information era, echo chambers formed on online social networks represent an environment conducive to group polarization through selective interaction of users with different people or information sources promoting a particular type of content. This describes how beliefs are amplified or reinforced through communication and repetition within a closed system (Ciampaglia, Menczer and TheConversationUs 2018). The entrenchment of misinformation is facilitated by exploiting cognitive biases, creating a self-reinforcing loop that perpetuates and amplifies misinformation and impedes critical thinking and informed discourse (Donis 2021).

Although the concept of the echo chamber is not a new phenomenon, and they do not only occur online, Eli Pariser also coined the term to describe the less-than-transparent process of data filtering by websites that "create a unique universe of information" customized for each individual (Pariser 2011, 10). One of the greatest dangers these filters pose is self-intoxication or indoctrination of users with their ideas through high consumption of familiar information material and exclusion of data that contradict individuals' beliefs (Pariser 2011, 13).

Misinformation often occurs in the echo chamber because these isolated environments provide fertile ground for deliberately disseminating false or misleading information to an audience less likely to question or critically examine it due to prevailing cognitive biases (Ceron and de-Lima-Santos 2023). According to a study by the American Psychological Association, misinformation campaigns manipulate cognitive biases to shape social perceptions by presenting information that aligns with people's existing beliefs, confirming their misconceptions and misjudgments. When false initial information (anchoring biases) is introduced, it becomes a reference point from which individuals find it challenging to move away. Furthermore, in the echo chamber, groupthink exacerbates the lack of critical control, as the unanimous group opinion suppresses dissent and encourages the acceptance of misinformation as truth (Nisbet and Kamenchuk 2019). Another Harvard Kennedy School Misinformation Review study found that misinformation campaigns can exploit cognitive biases such as confirmation bias, availability bias, and the illusory truth effect to manipulate public opinion (Murphy et al. 2023). In other words, in the echo chamber created in online environments, the seeds of misinformation can quickly take root by taking advantage of users' misguided beliefs, thereby validating their less-than-factual ideas.

Research

This research aims to investigate the extent to which social media algorithms facilitate the formation of echo chambers and how manipulative actors exploit these platforms to spread disinformation. The central hypothesis of this study posits

that social media algorithms play a significant role in creating echo chambers, which are then exploited by individuals or states for financial or political gain. These actors employ sophisticated techniques to use the cognitive biases of social groups, leading to a cycle where individuals seek affirmation and group identity and become complicit in propagating disinformation. Over time, these echo chambers are hypothesized to become self-sustaining as group members increasingly engage in selective information-seeking behaviors reinforcing their pre-existing beliefs. By identifying the mechanisms through which echo chambers are formed and exploited, this study aims to provide insights into how social media platforms can be redesigned to mitigate the spread of disinformation (Barberá 2020).

Purpose

The rise of extremist ideologies and the intensifying polarization within societies pose significant risks to the core values of democratic institutions, making this research highly relevant (Travierso 2021). The entrenchment of polarized viewpoints has been linked to the escalation of social tensions and the erosion of the common ground necessary for democratic discourse in contemporary society. By understanding the relationship between algorithm-driven content curation, echo chambers, and disinformation, this study aims to develop more robust counterstrategies to safeguard public discourse against the divisive effects of such phenomena (Travierso 2021).

The insights gained from this research are vital for policymakers, social media platforms, and civil society to mitigate the threats posed by disinformation to democratic processes, social cohesion, and the sanctity of factual debate. This research advances academic discussion in an era where the digital domain increasingly mirrors and influences the physical world. It is a critical resource for those endeavoring to preserve and enhance democratic engagement in the face of growing digital challenges (Travierso 2021).

Literature Review

The present study builds upon existing research on the echo chamber, misinformation, and cognitive bias to explore the manipulation of public opinion online through algorithmic personalization on social platforms. This study investigates Filter Bubble Theory (Pariser 2011) and the Law of Group Polarization (Sunstein 1999), as well as insights from "Networked Propaganda" (Benkler, Farris and Roberts 2018), to provide new perspectives on this topic.

Algorithmic personalization on social platforms reinforces echo chambers by creating feedback loops that perpetuate users' existing beliefs. This effect is further intensified by the design of online platform algorithms that aim to keep users

engaged. Ally Daskalopoulos et al. (2021) noted in the Detroit Regional Chamber Report that online news consumers often interact with fake news and spread unverified information, contributing to political polarization and misinformation. As a result, the echo chamber on digital platforms has significant implications for public discourse and societal polarization. They limit exposure to diverse viewpoints and undermine democratic processes (Garrett 2009; Woolley and Howard 2016; Lewandowsky, Ecker and Cook 2017).

While Axel Bruns (2019) suggests that users can interact with a broader range of content than Filter Bubble Theory assumes, indicating the existence of ideologically diverse social networks, most research emphasizes the role of selective exposure and confirmation bias in stimulating societal division into distinct groups (Cinelli et al. 2021).

The present study sheds light on how misinformation becomes ingrained in the human mind. Diaz Ruiz and Nilsson (2023) have shown that disinformation circulates through identity- and belief-based grievances. Their research focuses on the echo chamber formed on YouTube around the idea that the Earth is flat, illustrating how rhetoric plays a crucial role in spreading misinformation. Manipulative techniques appealing to beliefs and values related to cultural identity and religion are not the only ways to trap people in echo chambers with misinformation. Another study of US participants found that emotions generated by fake news headlines play a defining role in their redistribution and virality. Generally, study participants were more likely to give credence to news headlines aligning with their beliefs. At the same time, they were more tempted to dismiss news stories that elicited strong negative emotions and were not in line with their political values (Christy et al., 2021).

In sum, misinformation is constructed to reach the audience's emotional side, not only the cognitive component, taking advantage of the most used communication channels nowadays - the online environment.

The article examines the emotions influenced by misinformation, as per the research by Dag Wollebæk et al. (2019). It was found that fear and anger are the most powerful emotions associated with online misinformation, especially in the political context. The study surveyed Norway in 2017 and discovered that people who experience fear are more open to debate, both for and against their beliefs. This leads to an increase in the search for information, especially from opposing viewpoints, thereby improving the quality and quantity of information gathered. Fear and anxiety also prompt individuals to question specific facts and events. On the other hand, anger makes people rely on heuristic cues and existing routines, reducing the desire to seek new information. This leads to risk-seeking behavior and can worsen biased reasoning in the political arena, resulting in echo chambers.

The paper explores under-researched areas in the study of the echo chamber and misinformation, particularly in the Romanian social context. It highlights the need for more comprehensive research on how rapidly evolving topics in social media discourse influence user engagement and potentially lead to shifting beliefs

or values. The possibility of users experiencing cognitive shifts, especially in the context of misinformation, is noted as an area for further investigation. This includes examining how users may navigate between emotional and rational responses and factual and misleading information. The extent of research on these dynamics in the current literature warrants further exploration

Additionally, there is a phenomenon known as repetitive posting and the "flying monkey" effect, which has yet to receive adequate attention. This term, used metaphorically, describes the role of group members in echo chambers who contribute to amplifying and reinforcing misinformation. As a result, false narratives can go viral and persist despite their lack of accuracy.

Finally, there needs to be more understanding of the tangible impact of such digital phenomena on real-world scenarios, particularly regarding their negative consequences on Romanian society.

Data Collection

The present study aims to analyze a Facebook group in Romania, highlighting the phenomenon of extremism and division observed to be growing both in the countries of the European Union and globally. Online platforms, especially social media, have become fertile grounds for the spread of political extremism and polarization. This phenomenon is amplified by algorithms that personalize content according to user preferences, thus favoring the formation of echo chambers. In these spaces, individuals are isolated in informational bubbles, exposed only to information that reinforces their initial beliefs. According to organizations such as OpenDemocracy, in countries such as Romania, there has been an alarming increase in extremism and political polarization, a phenomenon that is becoming more and more visible in online echo chambers. This increase in polarization underscores a sharp ideological divide, reflected in the digital divide.

A relevant example of extremist tendencies in Romania is represented by an individual who has become notorious for their extremist rhetoric and behavior, especially during the pandemic. This person has expressed opposition to the COVID-19 restrictions and has taken a vocal anti-vaccination stance. With a strong presence on social media platforms like Facebook, where they actively participate in various groups or pages, this case highlights the manifestation of online extremism and its significant impact on Romanian society. The initial study explored Facebook groups that promoted perspectives associated with this person, with posts subsequently extracted through direct observation. The selection process aimed to identify the most significant and prominent groups, ultimately focusing on a single group with 93,800 members and an average of ten posts per day (the most relevant and current in terms of posts associated with this individual). Manual data collection took place over the course of a year, from January to December 2023, allowing for a detailed analysis of the group's

discourse and providing a deep understanding of the current situation, especially in the context of the anticipated elections in Romania in 2024.

Analyses methods

The study combines advanced computational methods such as **Latent Dirichlet Allocation (LDA), Natural Language Processing (NLP), Named Entity Recognition (NER), and machine learning** with manual analytical techniques to provide a comprehensive analysis of the echo chamber (Akhtar et al. 2023). The study employs a multifaceted approach to analyze the dynamics within an echo chamber, leveraging various methods to dissect the intricacies of language use, the spread of disinformation, and engagement patterns. The study begins with **topic modeling,** categorizing, and identifying the main themes and subjects in various social media posts. LDA or similar algorithms often drive this method and are crucial for understanding the dominant topics and narratives within the echo chamber. The study also includes a **feature extraction component,** where the mentions of specific people, places, and institutions are listed and analyzed. This is typically achieved through NER, a subfield of NLP, which helps identify and classify critical elements in the text.

Moreover, the study delves into the analysis of **emotional and manipulative language**. This involves examining the text for language that is designed to evoke emotional responses or manipulate perceptions, a process that can be enhanced by sentiment analysis algorithms and NLP tools (He, Hu and Pei 2023). Additionally, the study analyzes the **posts' frequency, engagement rates, and timing.** This involves statistical analysis and data visualization techniques to understand activity patterns and user engagement within the echo chamber.

In this study, I have developed a matrix for analyzing disinformation, drawing on extensive literature reviews, studies on disinformation, and my personal experience with multiple disinformation texts. It is important to note that while this matrix offers a structured framework for comprehensive and systematic examination, it could be better and remains a subject of ongoing research and optimization. The approach minimizes subjective interpretations by relying on predefined criteria, enhancing objectivity.

This structured approach is crucial for maintaining objectivity, as it relies on predefined criteria for evaluation, minimizing subjective interpretations. The consistency afforded by a standardized matrix is invaluable, as it allows for the replication of studies and comparison of findings, fundamental aspects of scholarly research. Moreover, disinformation is inherently multifaceted, and a matrix facilitates a multidimensional analysis, enabling researchers to delve into various elements such as language use, source credibility, and logical consistency. This comprehensive approach yields a deeper understanding of the complexities of disinformation.

Disinformation Criteria	Criteria Explanation
Emotionally and Sensationalist Language	Uses emotionally laden or sensational language to provoke emotional responses
Misrepresentation of Facts or Context	Presents facts inaccurately or out of context
Appeal to Fear and Urgency	It plays on fears and creates a sense of urgency.
Lack of Credible Evidence	Absence of backing from credible sources
Questionable Source Credibility	Relies on sources with dubious credibility
Factual Inconsistencies	Contradictions within the content or with facts
Contradiction of Established Understanding	Significantly deviates from widely accepted knowledge without substantial evidence.
Polarizing Us-vs-Them Rhetoric	Uses divisive language to create a sense of ingroup and outgroup
Overgeneralization and Stereotyping	Employs generalizations and stereotypes
Selective or Omitted Information	Leaves out critical information or selectively presents facts
Manipulation of Quotes or Sources	Alters quotes or sources to mislead
Lack of Transparency in Sourcing	Does not disclose sources, obscuring the origin of information
Inconsistent Logic or Argumentation	Arguments based on flawed reasoning or inconsistent logic
Use of Unverified or False Visuals	Relies on visuals that are unverified or proven false
Hyperbolic or Provocative Headlines	Uses headlines that are exaggerated or provocative
Opinion Presented as Fact	Presents opinions as if they were information
Anachronism	Refers to events or contexts inaccurately to mislead
Logical Fallacies	Employs reasoning that contains logical errors

Figure 1 Disinformation matrix

Findings

In the next phase, I conducted an in-depth analysis of the data extracted. This dataset included parameters like posting dates, message contents, user profiles, network contexts, likes, interaction rates, and unique identifiers. This comprehensive data collection enabled the identification of various narratives, hate speech instances, emotive language use, negative sentiments, and widespread amplification techniques within the echo chamber's discourse. However, while these parameters provide insights into the group's structure and dynamics, they might not directly reveal the nature of the narratives. To fully grasp the narratives, a qualitative examination of the message content is essential, looking at the themes, ideologies, and rhetorical strategies used in the posts.

a) *Creating a Sense of Community and Belonging* – I observed that using inclusive language was a powerful tool for fostering a sense of community and belonging. Phrases like "we," "together," and "us," along with hashtags such as #solidaritate (solidarity), #Împreună (together), #Comunitate (community), emphasize a shared identity among group members. Additionally, hashtags like #TradițiaNeDefinește (tradition defines us), #SchimbareaDeCareAvemNevoie (the change we need), #ErouAlNostru (our hero), #unire (unity), #RespectPentruCulturaNoastră (respect for our culture) further reinforce this sense of belonging and shared values. This language not only reinforces the feeling of being part of a distinct and cohesive group, appealing to those seeking

a sense of belonging but also plays a significant role in shaping social identity. Using such 'us/we' narratives taps into key psychological aspects like social identity, ingroup favoritism, and outgroup derogation. Social identity theory explains how an individual's self-concept is influenced by their membership in a social group. Ingroup favoritism, the preference for one's group, and outgroup derogation, the negative perception of other groups, are prevalent in such settings. In certain situations, spreading false information within an echo chamber can have severe consequences, such as exacerbating group polarization and reinforcing negative perceptions of outside groups. According to Sunstein (1999), social media platforms can often act as echo chambers, promoting a sense of group identity and encouraging users to seek information that aligns with their beliefs. This can lead to a lack of critical thinking and dismissing opposing viewpoints, ultimately exacerbating negative attitudes toward outside groups. This environment can also facilitate the spread of fake news and disinformation without being challenged.

b) *Growth and repetitive messages* – In the context of disinformation and the rapid expansion of echo chambers, the group I studied represents a relevant example. Established on October 25, 2021, it has consistently grown. In just the last week of the study period, the group expanded by 174 members, averaging a daily increase of around 130 members. This escalating growth rate indicates the group's burgeoning influence and reach. Significantly, within a single year, the group's posts elicited over 705,826 reactions (likes), reflecting a high level of engagement among its members. This pattern of rapid growth and increased engagement is characteristic of echo chambers, where repetitive messaging and targeted content can quickly attract and retain a growing audience, amplifying disinformation's spread and impact. Public activities, community outreach, and direct engagement in the physical world can also contribute to the group's online growth. The group can attract individuals who carry their interest and involvement into the online space by establishing a presence in the offline world through events, demonstrations, or local campaigns. This synergy between offline actions and online engagement helps expand the group's reach further and solidify its influence in the digital and the real worlds.

c) *Reinforcing a Common Identity* – Echo chambers often use language that strengthens the perception of a unified group identity, shifting focus from individuals to the collective. This common identity is built around shared beliefs, opinions, or ideologies, serving as a powerful unifying force within the group. For instance, the opinion leader talks about the group he represents and online supporters in a post like this: "our #patriotic spirit firmly stand in #war against the import attempts of foreign #customs that have no connection with our #national identity, (…)" clearly demonstrates this dynamic.

In this post, the assertion of national identity is evident, as it frames the situation as a "war" against foreign customs, emphasizing the importance of maintaining a

distinct Romanian identity. Although not explicitly stated in this specific portion, the overall message of the post advocates for celebrating Romanian traditions and rejecting foreign ones, reinforcing the idea that national customs are integral to the group's identity. The post positions the broadcaster as defender of national identity against foreign customs, creating a narrative of protection and resistance that strengthens group cohesion and unity. Using terms like "#war" and the emphasis on "our patriotic spirit" serves as a call to action for supporters to rally around the cause of preserving national identity, further solidifying the group's shared identity.

d) Reducing Receptivity to Opposing Viewpoints – In another post, the public figure states about another personality that she "*îi îndeamnă pe tineri să nu mai plece peste hotare: ‚Vă încurajez să vă proiectați un viitor în România*'" (Translation: „*urges young people not to go abroad: ‚I encourage you to design a future in Romania'*"), the comments of the group members predominantly indicate strong disapproval and cynicism compared to the encouragement of young people by another opinion former compared to the sympathetic one to stay in Romania. Many commenters view this as a disingenuous or manipulative gesture. This uniformity in sentiment and the lack of visible counterarguments or diverse perspectives suggest an echo chamber effect, where a single viewpoint predominates and dissenting opinions are either absent or dismissed. Some of the messages are:

- *Ristea D*: "Da stați în țara dragi tineri că au nevoie de carne de tun."* - *Translation: "Yes, stay in the country, dear young people, because they need cannon fodder."* This comment expresses cynicism towards young people staying in Romania, suggesting they are needed only as expendable resources, implying a lack of genuine concern for their well-being.
- *Monica P*: "Mars tradatorul! Vrei carne de tun?!"* - *Translation: "Go away, traitor! Do you want cannon fodder?!"* Echoing a similar sentiment, with this comment, the author labels the unloved public figure as a "traitor" and accuses him of wanting young people to be "cannon fodder", indicating a solid mistrust and negative perception.
- *Rodica I*: "Să îi trimiți în război nemernicule... Pupincuriști ai globaliștilor."* - *Translation: "Send them to war, you scoundrel... Bootlickers of the globalists."* This comment further intensifies the negative sentiment, accusing another public figure of sending young people to war and aligning himself with the globalists, reflecting deep-rooted mistrust and hostility.

When group identity becomes intertwined with specific beliefs or ideologies, any challenge to these beliefs feels like a personal attack on the group and, by extension, on the individual. This can result in a defensive stance against external information, further entrenching members in their existing beliefs. The strong sense of identity and belonging fostered by inclusive language leads to decreased openness to opposing viewpoints.

e) Frequent use of negative language – Utilizing the disinformation matrix analysis tool, I found that a significant portion of the language used in the echo chamber,

47,25% of the message, is harmful or manipulative. This high rate of negativity is crucial in echo chambers for several reasons. Firstly, it reinforces shared beliefs among group members, fostering a strong sense of unity against perceived outgroups or opposing ideas. This is often achieved by establishing a clear 'us versus them' dynamic. Secondly, negative language evokes stronger emotional reactions than neutral or positive language, leading to increased engagement within the echo chamber. This heightened emotional response can further entrench members' beliefs. Lastly, negativity in echo chambers contributes to the polarization and radicalization of opinions, as it increases members' resistance to outside information or alternative viewpoints. The significant presence of negative words, as revealed by our analysis, underscores the role of emotional and divisive language in shaping the dynamics within echo chambers.

f) Spontaneous Communication and Strong Emotion – In the analysis of the echo chamber's communication patterns, I encountered a notable use of manipulative language, with 2,076 instances identified across various posts. This manipulative language included words like: '*șocant (shocking)*,' '*brutală (brutal)*,' '*limitate (limited)*,' '*cenzură (censorship)*,' '*șocat (shocked)*,' '*devastatoare (devastating)*,' '*incendiară (incendiary)*,' '*limitărilor impuse (imposed limitations)*,' '*un eveniment pe care nu-l puteți rata (an event you can't miss)*,' '*mintit (lied)*,' '*manipulare totala (total manipulation)*,' '*arma (weapon)*,' '*profile false (false profiles)*,' '*stiri false (false news)*,' '*insultele (insults)*,' '*să nu mai îndrăzniți vreodată (never dare again)*,' '*haos (chaos)*,' '*tragedie (tragedy)*,' '*tristă (sad)*,' '*devastatoare (devastating)*,' '*suferă (suffer)*,' '*demisia (resignation)*,' '*esențială (essential)*,' '*vânzătorii (sellers)*,' '*gâtul nostru (our throat)*,' '*acuză (accuse)*' *and* '*ăștia (these)*'. These manipulative words and expressions, filled with emotional triggers, are frequently used within the group to shape opinions and emotions. A closer examination of 825 posts reveals that, on average, each contained approximately 5.78 manipulative words, demonstrating the group's reliance on persuasive language to shape opinions and emotions. The identified manipulative words, as seen in phrases like 'DEMITEREA LUI ARAFAT, ARESTAREA SI CONDAMNAREA LUI PE VIATA PT GENOCID' (Arafat's resignation, arrest, and life sentence for genocide), 'PENTRU A NU SPUNE ADEVĂRUL, MI-AU INTERZIS SA VORBESC' (To prevent me from telling the truth, they prohibited me from speaking), and 'MANIPULAREA DNA, FILMARI PLATITE LA COMANDA SI SCOASE DIN CONTEXT' (DNA manipulation, paid-for videos on demand taken out of context), are designed to evoke strong emotional reactions and shape opinions. This underscores the use of emotionally charged language to influence and manipulate readers or the audience. Additionally, communication within these hyper partisan echo chambers is characterized by spontaneity and a high degree of emotional intensity. This is exemplified by the frequent use of swear words and exclamation marks, which act as markers of strong emotions and heightened arousal states. In such a polarized environment, these expressions of emotion are often indicative of either strong agreement or vehement disagreement with the shared beliefs and

biases prevalent within the group.

Additionally, the presence of swear words in the group's discourse indicates a less formal yet more impassioned and, at times, aggressive communication style (Wollebæk et al. 2019). In hyperpartisan settings, such usage of swear words often serves dual purposes: expressing intense emotions like anger or disdain, especially towards opposing views or groups, and reinforcing ingroup unity against perceived outgroups. This linguistic feature contributes to the echo chamber's charged and often contentious atmosphere. Similarly, using exclamation marks can amplify the emotional intensity of a message. It can indicate excitement, surprise, anger, or urgency. In hyperpartisan echo chambers, exclamation marks can enhance the emotional resonance of messages, making them more impactful and memorable.

g) *A dynamic and strategic shifting of topics in the same echo chamber* (see Figure 1) – This shifting serves multiple purposes: firstly, it hooks subjects from diverse backgrounds and with varying interests, and secondly, it plays a crucial role in altering perceptions. The analyzed echo chamber can effectively manipulate the perceptions of its members by rapidly transitioning from one topic to another, thereby shifting their emotional and cognitive states. This tactic is insidious as it does not allow the brain enough time to adjust from processing emotional content to rational thought.

The discussions cover a wide range of topics that cater to the diverse interests and beliefs of the audience. Categories such as Social Issues, Romanian Culture and Identity, Media and Communication, Politics, Nationalism, Religion, and Lifestyle serve as practical tools to keep members engaged and active within the echo chamber. However, it is essential to note the manipulative technique, where a neutral fact is validated before subtly transitioning into disinformation and hate speech. This seamless shift makes it difficult for members to discern the change in content integrity, resulting in an emotional response to manipulated or false data instead of a rational engagement with genuine information.

This tactic of topic shifting and blending information with disinformation keeps the members engaged and effectively alters their perceptions. Studies, such as one conducted by the Reuters Institute for the Study of Journalism, have shown that the constant oscillation between different topics and between emotional and rational states undermines the ability of individuals to critically evaluate the information being presented, leading to a gradual acceptance of disinformation as truth (Brennen 2019). This phenomenon highlights echo chambers' sophisticated strategies to manipulate opinions and beliefs. Another study conducted by the Harvard Kennedy School found that most regular internet users globally worry about misinformation, with young and low-income groups expressing the highest levels of concern (Knuutila, Neudert and Howard 2022).

The constant oscillation between different topics and between emotional and rational states undermines the ability of individuals to critically evaluate the information being presented, leading to a gradual acceptance of disinformation

as truth. This phenomenon highlights echo chambers' sophisticated strategies to manipulate opinions and beliefs.

Categories	Hashtags categorization – NLP modeling
Social Issues	#CopiiiSuntViitorul (Children are the future), #oameni (people), #parinti (parents), #Comunitate (community), #gradinita (kindergarten), #scoala (school), #sociale (social), #Împreună (together), #solidaritate (solidarity), #binele (goodness), #respect (respect), #drepturile (rights), #justitie (justice), #doaradevarul (only the truth), #drepturi (freedoms), #amintire (memory), #înţelegere (understanding), #recunoştinţă (gratitude), #adevar (truth), #toleranţă (tolerance), #hotărâţi (determined), #speciale (special), #TraficulDeDroguri (drug trafficking), #ConfidenţialitateFinanciară (financial privacy), #LibertateEconomică (economic freedom), #SchimbareaDeCareAvemNevoie (the change we need), #pace (peace), #incompetenta (incompetence), #nurazboi (no war).
Romanian Culture and Identity	#decembrie (December), #TradiţiaNeDefineşte (Tradition defines us), #Datini (Customs), #istoric (historical), #OraşulCuInimăVeche (The city with an old heart), #tebea.
Media and Communication	#DezvaluiriCuImpact (Impactful revelations), #tiktok, #televiziuni (television), #socialmedia, #stiri (news), #facebook, #televiziune (television), #tipografie (printing), #dezvaluiri (revelations), #ziare (newspapers), #informatie (information), #actual (current), #live, #instagram, #vocea (voice), #dezbatare (debate), #Laudăţium (praises).
Politics	#SchimbareaDeCareAvemNevoie (The change we need), #pace (peace), #incompetenta (incompetence), #nurazboi (no war), #sieusuntvoluntar (I am a volunteer), #DemisieArafat (Arafat resignation), #raedarafata (Arafat's resignation), #ResponsabilitateSauDemisie (Responsibility or Resignation), #uk (UK), #canada (Canada), #germania (Germany).
Nationalism	#Conştienţă (consciousness), #bucovina (Bucovina), #RomâniaReală (Real Romania), #patriotism (patriotism), #tara (country), #naţiune (nation), #patriot (patriot), #bunderomânia, #ErouAlNostru (Our hero), #roman (Romanian), #romăne (Romanian), #demnitate (dignity), #unirea (union), #istorie (history), #distrugă (destroy), #romania (Romania), #unire (union), #independent (independent).
Religion	#mihail (Michael), #sfint (saint), #ÎngeriiPăzitori (Guardian Angels), #cer (heaven).
Lifestyle	#iubire (love), #poet (poet), #carte (book), #simpozion (symposium), #viata (life), #cantece (songs), #dramaturg (playwright), #viitorul (future), #MoştenireLiterară (literary heritage), #RespectPentruCulturaNoastră (respect for our culture).

Figure 2   Rapid shifting topics among multiple hate and conspiracy messages

h) *Utilizes a broad spectrum of negative themes to engage and influence its members* – the wide range of categories used in the echo chamber's discourse reflects a sophisticated strategy to engage members across various emotional and intellectual levels. This strategy strengthens the echo chamber's internal cohesion and impacts broader societal dynamics, contributing to polarization, misinformation, and the erosion of healthy public discourse.

The prevalence of categories like *Negative Connotations and Slander*, *Discrimination and Bigotry*, and *Oppression and Coercion* indicates a deliberate strategy to invoke strong emotional reactions, often negative, to solidify group identity and beliefs. The presence of categories like *Solutions and Innovations*, *Support and Solidarity*, and *Hope and Aspiration* indicates that the echo chamber also uses positive and constructive themes. This dual approach of combining negative and positive narratives maintains engagement and bolsters the group's cohesive narrative. Moreover, categories such as *Legal and Ethical Issues*, *Social*

and Cultural Dynamics, and *Environmental and Ecological Awareness* show that the echo chamber does not shy away from complex and multifaceted issues. However, discussing these issues often needs more nuance and is framed within the echo chamber's ideological perspective.
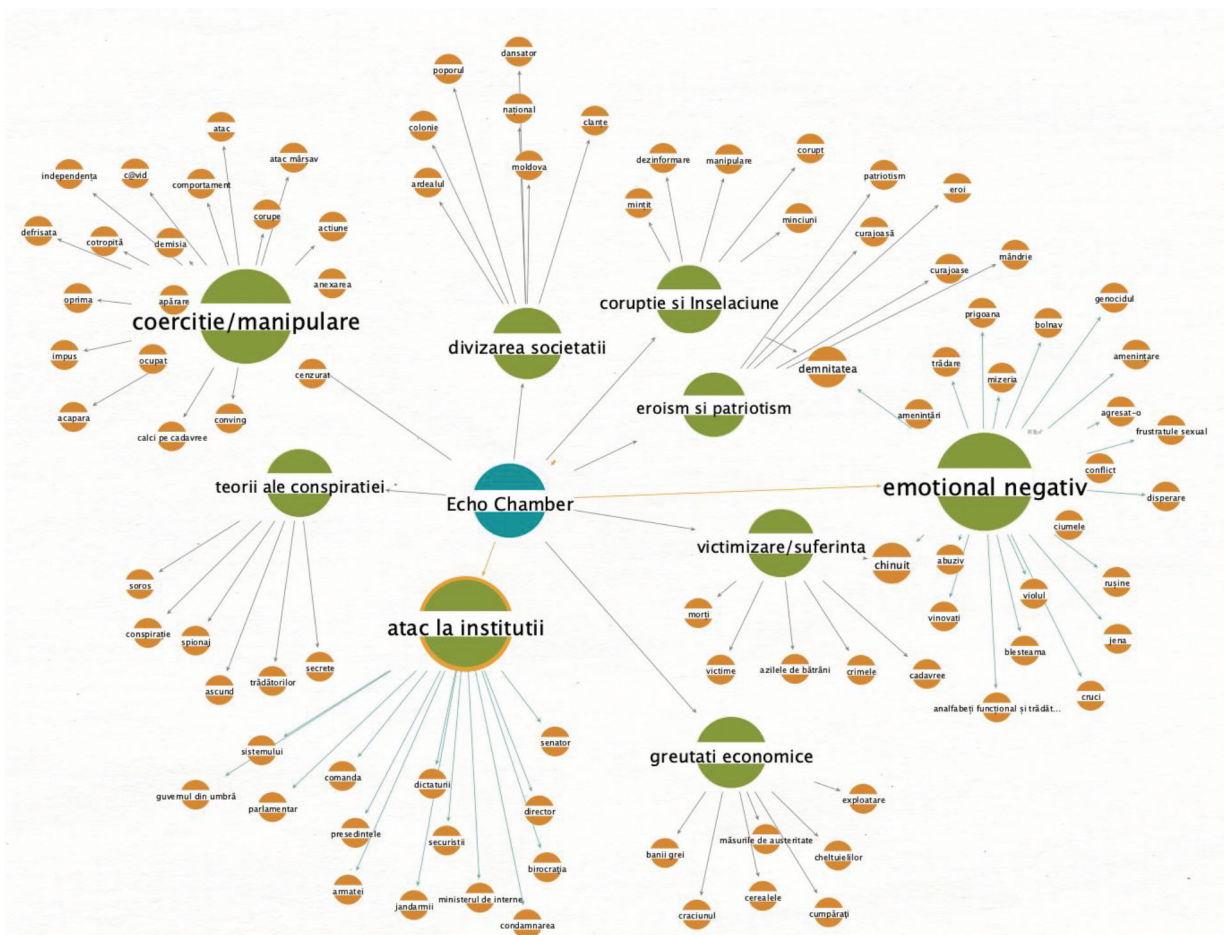
Figure 3   Topic modeling and clusterization on hateful, emotional, or negative content keywords found in the echo chamber

The keywords and topics in the echo chamber (see Figure 3) message suggest a complex and multifaceted approach to spreading disinformation and manipulating public opinion. Each set of keywords aligns with different aspects of disinformation strategies, aiming to influence perceptions, deepen societal divisions, undermine trust in institutions, and provoke emotional responses:

➤ **Coercion/Manipulation**: Keywords like "conving" (convincing), "calci pe cadavre" (stepping on corpses), "a acapara" (hoarding), "ocupa" (occupy), "impus" (impose), "oprima" (oppress), "defrişată" (deforestation), "cotropită" (invaded), "demisia" (resignation), "independenţa" (independence), "COVID" (COVID), "comportament" (behavior), "corupe" (corrupt), "atac" (attack), "atac mârşav" (malicious attack), "acţiune" (action), "anexare" (annexation), "cenzurat" (censored) suggest a narrative aimed at portraying forceful or deceitful actions,

possibly attributing them to specific groups or individuals to discredit them or justify aggressive or unethical behaviors.

➢ **Dividing Society**: Keywords such as "Ardealul" (Transylvania), "colonie" (colony), "poporul" (people), "trădător" (traitor), "Moldova" (Moldova) are likely used to exacerbate regional, ethnic, or national tensions, suggesting an effort to fragment societal unity by highlighting or fabricating divisions and fostering resentment among different communities.

➢ **Corruption and Deception**: Words like "mințit" (lied), "dezinformare" (disinformation), "manipulare" (manipulation), "corupție minciuni" (corruption lies) are typical in narratives aimed at undermining trust in individuals, organizations, or processes by accusing them of dishonest or fraudulent behavior, often without evidence.

➢ **Conspiracy Theories**: The mention of "conspirație" (conspiracy), "secrete" (secrets), "trădătorilor" (traitors), "ascund" (hide), "spionaj" (espionage), "Soros" (Soros) indicates the promotion of unfounded theories that claim shadowy forces are orchestrating events behind the scenes. These narratives often rely on prejudice, paranoia, or speculative connections to explain complex situations through simplified, misleading lenses.

➢ **Attacks on Institutions**: Keywords targeting "sistemului" (the system), "guvernul din umbră" (shadow government), "parlament" (parliament), "comandă" (command), "președintele" (president), "armatei" (army), "dictaturii" (dictatorship), "jandarmii" (police), "securiștii" (secret police), "ministrul de interne" (interior minister), "condamnare" (conviction), "birocrație" (bureaucracy), "director" (director), "senator" (senator) suggest an effort to erode confidence in governmental and societal institutions by portraying them as corrupt, oppressive, or illegitimate.

➢ **Economic Hardships**: Terms like "banii grei" (big money), "Crăciunul" (Christmas), "cerealele" (grains), "măsurile de austeritate" (austerity measures), "cumpărat" (bought), "cheltuielilor" (expenses), "exploatare" (exploitation) reflect concerns about economic conditions and policies, possibly twisting facts or context to incite anger or despair about the economic situation, attributing blame to specific entities or policies without a balanced perspective.

➢ **Victimization/Suffering**: With keywords such as "morți" (deaths), "victime" (victims), "azilele de bătrâni" (nursing homes), "crimele" (crimes), "cadavre" (corpses) the narrative focuses on highlighting real or imagined instances of suffering to elicit empathy, anger, or fear, often to sway opinion or justify radical viewpoints or actions.

➢ **Negative Emotional Appeals**: Phrases like "chinuit" (tortured), "abuziv" (abusive), "vinovați" (guilty), "analfabeți funcționali și trădători" (functional illiterates and traitors), "blestem" (curse), "viol" (rape), "cruci" (crosses), "rușine" (shame), "ciumele" (plagues), "disperare" (desperation), "conflict" (conflict), "frustrat sexual" (sexual frustration), "agresor" (aggressor), "amenințare" (threat), "genocid" (genocide), "bolnav" (sick), "mizerie" (filth), "prigoană" (persecution), "trădare" (betrayal) are designed to evoke strong negative emotions, aiming to

manipulate the audience's feelings to provoke outrage, fear, or hatred, bypassing rational analysis.

The goal is often to polarize society, distract from substantive issues, or consolidate power by creating an environment where rational discourse is overshadowed by fear, suspicion, and anger.

Discussion

Examining this echo chamber on Facebook yields valuable insights when contextualized within the cognitive biases and disinformation framework, thereby expanding our understanding of the phenomenon at hand. My observations were meticulously conducted by analyzing social categories and topic clustering inside the Facebook group. The diverse array of thematic orientations and manipulative tactics prevalent in the echo chamber is congruent with and leverages diverse cognitive biases. Biases like confirmation bias, where individuals favor information that confirms their pre-existing beliefs, are particularly evident. The echo chamber's targeted content reinforces these biases, continually exposing members to viewpoints that align with their own, thereby perpetuating a cycle of reinforced beliefs and misconceptions.

Echo chambers act as effective incubators of disinformation by creating an environment where only specific perspectives are presented and opposing views are omitted or discredited. These chambers create a fertile ground for disinformation to flourish. The emotional and manipulative language used further exacerbates this issue, as it is designed to evoke strong reactions and discourage critical thinking, making members more susceptible to accepting false information as truth.

The use of identity-based themes like nationalism and patriotism, coupled with negative portrayals of outgroups, contributes to group polarization. This polarization is compounded by groupthink, a phenomenon where the desire for harmony or conformity within the group results in irrational or dysfunctional decision-making. In such an environment, questioning or challenging the group's beliefs becomes increasingly complex, leading to a more homogeneous and extreme set of views.

Implications

The findings from our study on one possible echo chamber on the social Facebook platform have profound implications for understanding the nature of echo chambers more broadly and devising strategies to combat disinformation. The mechanisms identified – such as the use of emotionally charged and manipulative language, the formation of a strong sense of community, and the presence of spontaneous, emotionally intense communication – are not unique to the groups we studied. The

dynamics we observed reflect a broader, worldwide pattern where echo chambers, marked by manipulative techniques, are emerging across diverse contexts, highlighting the need for a global approach to address these challenges. These environments are particularly conducive to extremism, providing a platform where extreme ideas are amplified and normalized, potentially leading to the radicalization of individuals or masses, especially within vulnerable populations.

Moreover, the influence of echo chambers extends to democratic processes and societal stability, polarizing public opinion, eroding shared truths, and undermining healthy democratic discourse. This can result in real-world destabilization, driven by actions based on distorted perceptions and misinformation. Addressing these challenges requires understanding the mechanics of echo chambers to develop effective strategies for combating disinformation. This involves raising awareness, promoting media literacy, encouraging exposure to diverse viewpoints, and addressing the structural aspects of social media that contribute to forming these chambers.

Additionally, the implications for policymakers and regulators are significant. They need to be cognizant of the effects of echo chambers on social cohesion and democratic integrity, possibly requiring a reassessment of social media's role in public discourse, implementing regulations to curb disinformation, and supporting initiatives that foster critical thinking and fact-checking.

Limitations

It is essential to acknowledge some of the study's limitations by noting several vital constraints that may impact the generalizability and scope of the findings. First and foremost, the study covers only a tiny percentage of the potential echo chambers, focusing specifically on specific social segments within Romania. Given the unique cultural and historical context of Romania, the strategies adopted by the owners of these echo chambers and the nature of the discourse within them may not be fully representative of similar phenomena in other countries or cultures (Arguedas et al. 2022).

Furthermore, cognitive biases, which play a significant role in forming and sustaining echo chambers, can vary considerably from one population to another. These biases are influenced by a multitude of factors, including cultural background, social context, and historic periods, which means that the cognitive patterns observed in the study may not be universally applicable. Additionally, while the research provides valuable insights into the workings of echo chambers and their impact, it needs to be more comprehensive. Several external factors that might influence the dynamics within echo chambers, such as political restrictions, economic conditions, or external events like wars in neighboring countries, were not within the scope of the analysis. These factors can have a significant bearing on public opinion and the spread of disinformation but were not accounted for in the study.

In summary, while the study contributes essential findings to understanding echo chambers in a specific social and cultural context, these limitations should be considered when extrapolating the results to other contexts or broader populations. Future research could address these limitations by incorporating a more comprehensive range of echo chambers across different cultural and social contexts and considering various influencing factors.

Conclusions

The findings from this study on a potential echo chamber within the Facebook platform offer profound insights into the nature of such environments and suggest strategies to counter disinformation. This research indicates that the mechanisms identified, including emotionally charged language, cultivating a strong community sense, and emotionally intense communication, likely reflect a broader global trend in echo chambers.

These echo chambers provide fertile ground for extremism, amplifying extreme ideas and potentially leading to the radicalization of individuals or groups, particularly within vulnerable populations. Furthermore, they can significantly impact democratic processes and societal stability by polarizing public opinion and eroding shared truths, potentially leading to real-world destabilization fueled by misinformation. Addressing these issues necessitates understanding echo chamber mechanics and a comprehensive approach involving media literacy, exposure to diverse viewpoints, and structural changes in social media. Policymakers and regulators also play a crucial role and should be informed of these implications to develop effective policies and regulations.

References

Akhtar, P., A.M. Ghouri, H.U.R. Khan, M.A.U. Haq, U. Awan, N. Zahoor, Z. Khan and A. Ashraf. 2023. "Detecting fake news and disinformation using artificial intelligence and machine learning to avoid supply chain disruptions." *Annals of Operations Research* 327: 633-657. https://doi.org/10.1007/s10479-022-05015-5.

Arguedas, A.R., C.T. Robertson, R. Fletcher and R.K. Nielsen. 2022. "Echo chambers, filter bubbles, and polarisation: a literature review." *Reuters Institute.* doi:10.60625/risj-etxj-7k60.

Barberá, P. 2020. „Social Media, Echo Chambers and Political Polarization." In *Social media and democracy. The State of the Field, Prospects for Reform*, editors Nathaniel Persily and Joshua A. Tucker, 34-55. https://www.cambridge.org/core/books/social-media-and-democracy/social-media-echo-chambers-and-political-polarization/333A5B4DE1B67EFF7876261118CCFE19.

Benkler, Y., R. Farris and H. Roberts. 2018. *Network Propaganda.* Oxford University Press. https://doi.org/10.1093/oso/9780190923624.001.0001.

Braddock, K. 2022. "Vaccinating against hate: Using attitudinal inoculation to confer resistance to persuasion by extremist propaganda." *Terrorism and Political Violence* 34 (2): 240-262. https://doi.org/10.1080/09546553.2019.1693370.

Brennen, J.S. 2019. "Misinformation: The evidence on its scope, how we encounter it, and our perceptions of it." *Reuters Institute for the Study of Journalism.* https://reutersinstitute.politics.ox.ac.uk/news/misinformation-evidence-its-scope-how-we-encounter-it-and-our-perceptions-it.

Bruns, A. 2019. "Filter bubble." *Internet Policy Review* 8 (4). https://doi.org/10.14763/2019.4.1426.

Ceron, Wilson and Mathias-Felipe de-Lima-Santos. 2023. "Disinformation Echo Chambers on Facebook." In *Fighting Fake Facts*, 61-90. MDPI Books. https://doi.org/10.3390/books978-3-0365-1347-8-4.

Ciampaglia, G.L., F. Menczer and TheConversationUs. 2018. "Biases Make People Vulnerable to Misinformation Spread by Social Media." *Scientific American.* https://www.scientificamerican.com/article/biases-make-people-vulnerable-to-misinformation-spread-by-social-media/.

Cinelli, Matteo, Gianmarco De Francisci Morales, Alessandro Galeazzi and Michele Starnini. 2021. "The echo chamber effect on social media." *Proceedings of the National Academy of Sciences* 118 (9): e2023301118. https://doi.org/10.1073/pnas.2023301118.

Daskalopoulos, A., N. Hernandez, F. Jason, H. Jenvey, D. Gustafson, R. Mosley, C. Rodriguez, N. Schoonover and S. Townsend. 2021. "Thinking outside the bubble: Addressing polarization and disinformation on social media." *CSJS Journalism Bootcamp.* https://journalism.csis.org/thinking-outside-the-bubble-addressing-polarization-and-disinformation-on-social-media/.

Donis, L. 2021. „How to filter bubbles and echo chambers reinforce negative beliefs and spread misinformation through social media." *Debating Communities and Networks XII.* https://networkconference.netstudies.org/2021/2021/04/25/how-filter-bubbles-and-echo-chambers-reinforce-negative-beliefs-and-spread-misinformation-through-social-media/.

Garrett, R.K. 2009. "Echo chambers online?: Politically motivated selective exposure among Internet news users." *Journal of Computer-Mediated Communication* 14 (2): 265-285. https://doi.org/10.1111/j.1083-6101.2009.01440.x.

He, L., S. Hu and A. Pei. 2023. "Debunking Disinformation: Revolutionizing Truth with NLP in Fake News Detection." https://arxiv.org/abs/2308.16328.

Horner, G.C., D. Galletta, J. Crawford and A. Shirsat. 2021. "Emotions: The Unexplored Fuel of Fake News on Social Media." *Journal of Management Information Systems* 38 (4): 1039-1066. doi:10.1080/07421222.2021.1990610.

Knuutila, A., L.M. Neudert and P.N. Howard. 2022. "Who is afraid of fake news? Modeling risk perceptions of misinformation in 142 countries." *Misinformation Review, Harvard Kennedy School.* https://misinforeview.hks.harvard.edu/article/who-is-afraid-of-fake-news-modeling-risk-perceptions-of-misinformation-in-142-countries/.

Lewandowsky, S., U.K.H. Ecker and J. Cook. 2017. "Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era." *Journal of Applied Research in Memory and Cognition* 6 (4): 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008.

Murphy, G., C. De Saint Laurent, M. Reynolds, O. Aftab, K. Hegarty, Y. Sun and C.M. Greene. 2023. "What do we study when we study misinformation? A scoping review of experimental research (2016-2022)." *Harvard Kennedy School Misinformation Review.* https://misinforeview.hks.harvard.edu/article/what-do-we-study-when-we-study-misinformation-a-scoping-review-of-experimental-research-2016-2022/.

Nikolov, D., D.F.M. Oliveira, A A. Flammini and F. Menczer. 2015. "Measuring online social bubbles." *PeerJ Computer Science.* https://doi.org/10.7717/peerj-cs.38.

Nisbet, E.C. and O. Kamenchuk. 2019. "The Psychology of State-Sponsored Disinformation Campaigns and Implications for Public Diplomacy." *The Hague Journal of Diplomacy* 14 (2): 65-82. https://doi.org/10.1163/1871191X-11411019.

O'Shaughnessy, N. 2020. „From disinformation to fake news: forwards into the past." În *The SAGE Handbook of Propaganda*, pp. 55-70. SAGE Publications Ltd. https://doi.org/10.4135/9781526477170.

Pariser, E. 2011. *The Filter Bubble: What the Internet Is Hiding from You.* https://dl.acm.org/doi/10.5555/2029079.

Ruiz, C. Diaz and T. Nilsson. 2023. "Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies." *Journal of Public Policy & Marketing* 42 (1): 18-35. https://doi.org/10.1177/074391562.

Sunstein, C.R. 1999. "The Law of Group Polarization." (John M. Olin Program in L. & Econ. Working Paper No. 91). https://dash.harvard.edu/bitstream/handle/1/13030952/The%20Law%20of%20Group%20Polarization.pdf.

Travierso, M. 2021. "Measuring magnetism: how social media creates echo chambers." *Nature Italy.* https://www.nature.com/articles/d43978-021-00019-4.

Vosoughi, S. and S. Arala. 2018. "The spread of true and false news online." *Science* 359 (6380): 1146-1151. https://doi.org/10.1126/science.aap9559.

Walter, S., M. Brüggemann and S. Engesser. 2018. "Echo Chambers of Denial: Explaining User Comments on Climate Change." *Environmental Communication* 12 (2): 204-217. https://doi.org/10.1080/17524032.2017.1394893.

Wollebæk, D., R. Karlsen, K. Steen-Johnsen and B. Enjolras. 2019. "Anger, Fear, and Echo Chambers: The Emotional Basis for Online Behavior." *Social Media + Society* 5 (2). https://doi.org/10.1177/2056305119829859.

Woolley, S.C. and P. Howard. 2016. "Automation, algorithms, and politics/ Political communication, computational propaganda, and autonomous agents – Introduction." *International Journal of Communication* (10). https://ijoc.org/index.php/ijoc/article/view/6298/1809.